INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

Running Head: EVALUATION OF THE GATB - COMPUTERIZED ADMINISTRATION

Evaluation of the General Aptitude Test Battery- Computerized Administration (GATB-CA)

Kevin C. Yeasting ©

Lakehead University

A Thesis submitted to the

Department of Psychology

in Partial Fulfillment of the Degree of

Master of Arts

August 1996

Acknowledgements

I would like to take this time to say a simple thanks to all of those who have contributed their efforts and ideas over the course of this project. Dr. C. Netley for his guidance and support, Dr. D. Mazmanian for his advice and assistance, and Walter Martin for allowing me to use the newly developed GATB-CA in this study. I would also like to thank my family for their continual support every step of the way. Finally, I would especially like to thank my lifelong companion, Sandy Jung, whose energy, encouragement, and passion for life has kept me moving forward always reaching for the stars. Never again will I have to dream alone.

Abstract

The General Aptitude Test Battery (GATB), a vocational test developed in 1947 by the United States Employment Service (USES), is one of the most widely used and researched instruments worldwide. A recent development important in improving the GATB was the introduction of a new computer administered version, the GATB-CA or GATB-Computerized Administration. Presently, there has been no research conducted with the GATB-CA. Thus, the purpose of the present study was twofold: (1) to compare the GATB-CA to the original GATB to test for equivalence in subject test scores and item response speed (measured by the total number of items completed) and (2) to investigate if scores from either of the GATB formats were able to predict success in subjects obtained from an academic program of study. Subjects consisted of 62 undergraduate students. The research involved a random assignment counterbalanced design with all subjects completing both the conventional and computerized GATB versions. Results showed that certain parts of the GATB-CA were equivalent to the original GATB (subtests 6 and 7; aptitudes P and S) while other portions were significantly different (subtests 1 and 4; aptitudes G, Q, and V). Despite these differences, the GATB-CA was able to predict academic success with essentially the same level of confidence as the original GATB. Thus, the GATB-CA was found to be closely related to but not equivalent to the original GATB.

Table of Contents

List of Tables

Evaluation of the General Aptitude Test Battery-Computerized Administration (GATB-CA)

The area of vocational and career guidance in psychology has enjoyed a period of prosperous growth in the last few decades. One particular instrument, the General Aptitude Test Battery (GATB), has stimulated a plethora of research in this area. After its introduction in 1947 by the United States Employment Service (USES), it rapidly accumulated an extensive research base and came to be recognized as "the best validated multiple aptitude test battery in existence for use in vocational guidance" (Manual for the GATB, Section I, 1986, p. iii).

In an era of increasingly advanced computer technology, it was inevitable that a computerized version of the General Aptitude Test Battery (GATB-CA) would be produced. The aim of the present study was to evaluate this new format of the GATB and compare it to the conventional GATB. As there has been no previous research using the GATB-CA, this study was a preliminary evaluation of the instrument.

Conventional General Aptitude Test Battery (GATB)

The General Aptitude Test Battery (GATB) was originally produced in 1947 by the U.S. Employment Service (Manual for the GATB, Sections I-IV, 1980; 1983a; 1983b; 1986). The initial intent was to:

> ...isolate and identify the basic aptitudes underlying the large number of aptitude tests then used by the Employment Service, and to select those few tests providing the best measures of these basic aptitudes for combination into a test battery particularly suitable for use in counseling (Manual for the GATB, Section III, 1980, p. 7).

The GATB is primarily used for vocational counseling and employment selection. The reason for this is that the performance profile an individual achieves on the test provides useful information regarding one's probable success in various occupations and also which occupations appear to be most suited to an individual's pattern of aptitudes.

Since its introduction, the GATB has remained relatively unchanged. The test consists of 12 timed subtests that measure 9 aptitudes. The 9 aptitudes are: 1) Intelligence - general learning ability, 2) Verbal Aptitude - ability to understand word and language meaning, 3) Numerical Aptitude - ability to perform arithmetic operations, 4) Spatial Aptitude - ability to comprehend two-dimensional representations of three-dimensional objects, 5) Form Perception - ability to perceive detail in objects or pictorial material, 6) Clerical Perception - ability to perceive detail in verbal or tabular material, 7) Motor Coordination - ability to coordinate eyes and hands/fingers rapidly and accurately in making precise movements, 8) Finger Dexterity - ability to move fingers and manipulate small objects with the fingers, and 9) Manual Dexterity - ability to move the hands easily and skillfully (see Manual for the GATB, Section II, 1983 for more detail). The 12 subtests consist of eight paper-and-pencil tasks and four manual performance tasks that involve apparatus manipulation (see Table 1).

GATB administration requires approximately two and a half hours and individuals examined by it should have a minimum level of grade 6 education. The GATB uses a system whereby raw scores for each subtest are converted to standard aptitude scores which have a mean of 100 and a standard deviation of 20. Originally, the development of GATB norms were based on the first 519 employed workers tested. In 1952, the GATB was again normed but this time with a larger

Table 1

Structure of the GATB

Factor	Aptitude	Subtests
I. Cognitive	G - Intelligence	Part 3 - 3-Dimensional Space
		Part 6 - Arithmetic Reasoning
		Part 4 - Vocabulary
	V - Verbal Aptitude	Part 4 - Vocabulary
	N - Numerical Aptitude	Part 2 - Computation
		Part 6 - Arithmetic Reasoning
II. Perceptual	S - Spatial Aptitude	Part 3 - 3-Dimensional Space
	P - Form Perception	Part 5 - Tool Matching
		Part 7 - Form Matching
	Q - Clerical Perception	Part 1 - Name Comparison
III. Psychomotor	K - Motor Coordination	Part 8 - Mark Making
	F - Finger Dexterity	Part 11 - Assemble
		Part 12 - Disassemble
	M - Manual Dexterity	Part 9 - Place
		Part 10 - Turn

sample of 4,000 workers between the ages of 18 and 54 years. This sample was "stratified to obtain proportional occupational representation of the general working population" (Manual for the GATB, Section III, 1980, p. 19). Presently, GATB norms have been established for a vast number of professional fields, some of which

are engineering, dentistry, nursing, teaching, business administration, accounting, marketing, and education (Dolke & Sharma, 1975).

The validity of the GATB has been extensively researched throughout the years. In an early review, Bemis (1968) summarized 424 studies involving over 25,000 subjects and concluded that the GATB was reasonably valid. In another review by Kujoth (1973), it was concluded that the validity of the GATB has generally been considered excellent and is supported by numerous and often highly specific studies. In addition to these early reviews, more recent studies have emerged to support the validity of the GATB. First, in 1989, the National Research Council released a report specifically addressing validity issues of the GATB (cited in Baydoun & Neuman, 1992). This report cited studies from over 750 criterion-based validity studies. Using meta-analysis techniques, the committee found the overall validity for the GATB to be approximately .30. This moderate correlation led the committee to state that the GATB is providing users with valuable information but this information should not be used as the sole determinant for employment selection. In addition, the committee also found that the GATB appeared to have moderate validity for a wide variety of different types of jobs. Second, Baydoun and Neuman (1992) assessed the future of the GATB and concluded that although the overall validity of the test is moderate, the usefulness of the battery should not be in question. Finally, to address concerns that there may be selection bias in the database of GATB validity studies, Vevea, Clements, and Hedges (1993) analyzed the 755 studies of validity completed on the GATB since 1947. They concluded that there was no support for selection effects producing a significant or substantial inflation of estimates of GATB validity.

With regards to investigating the reliability of the GATB, there have definitely been fewer studies. Overall though, it can be shown that the GATB is reliable when high school seniors or more highly educated subjects are retested within a three year span. In general, the reliability ranges from .80 to .90 on subtest and composite scores (Manual for the GATB, Section III, 1980).

Finally, the GATB has been shown to correlate quite highly with other more recently developed multiple-abilities batteries. For example, Hakstian and Bennet (1978) compared the GATB with the Comprehensive Ability Battery (CAB), which was developed in 1975, and the Differential Aptitude Tests (DAT), introduced in the late 1940s. They found that the GATB correlated highly with these two instruments. Furthermore, a recent study (Stoelting, 1990) compared the GATB with the Microcomputer Evaluation and Screening Assessment (MESA) produced in 1982. Findings showed that strong correlations existed between scores of the similarly named GATB and MESA aptitudes (often at the .01 level of significance). In particular, the GATB G and the MESA G, which both measure general intelligence, showed a significant correlation of -.64 (this inverse relationship occurs because in the GATB higher scores represent greater aptitude levels whereas in the MESA lower scores are indicative of greater aptitude levels).

To summarize this brief introduction to the conventional GATB, it is fair to say that the test is a simple yet robust assessment instrument that provides assessors with useful information. Many researchers familiar with the GATB have even referred to it as the best validated aptitude test battery ever developed (Janikowski, Berven, & Bordieri, 1991).

International Uses of the GATB

As previously mentioned, the GATB was primarily developed with job placement and occupational counseling as its central focus (U.S. Employment Service, Division of Testing Staff, 1978). It has been used extensively since 1947 by the U.S. Employment Service offices and since 1966 by the Canada Employment and Immigration Commission (Manual for the GATB, Section II, 1983). In addition, it has also been utilized in schools, unions, vocational rehabilitation centers, and various other authorized agencies (Baydoun & Neuman, 1992).

The U.S. Employment Service (USES) has also maintained a continuing test research programme. This programme has conducted studies on maturation of aptitude scores during high school years, test reliability, validation against training and academic success, development of Occupational Aptitude Patterns for use in counseling, effects of training and cultural exposure on test scores, and minority/non-minority comparisons of test validity (Droege, 1984). Today, literature on the GATB produced by the USES continues to grow and is now part of an occupational guidance system that links performance on the test with 348 subgroups that account for some 12,000 occupations (Dagenais, 1990).

Outside of North America, the GATB also carries the interests of other countries worldwide. The U.S. Employment Service has provided many individuals/organizations in other countries with specimen sets of the GATB. Follow-up data from USES report that the list of active and current users of the GATB now totals 68 users in 35 countries (Droege, 1984). Moreover, the review by Droege (1984) indicates that the GATB has been translated into many languages: French, German, Portuguese, Spanish, Italian, Hebrew, Japanese, Chinese, Korean, Dutch, Arabic, and Indonesian.

Much of the GATB research outside of Canada and the U.S. has been involved in the standardization and validation of the instrument with diverse populations. For instance, Howe (1975) sought to compare performance of U.S. and Australian subjects on the GATB and also to see if GATB aptitude measures correlate with commonly used Australian tests measuring similar skills. Results from the data of 2,917 subjects (1,355 Australian; 1,562 American) indicate that in most respects performance on the GATB was equivalent. Furthermore, results on the GATB measures G, V, N, and Q correlated highly with results obtained on similar Australian measures having much longer testing times. Similarly, another study conducted in Saudi Arabia (Dagenais, 1990) found that in comparing American and Saudi Arabian populations, the patterns of mean test scores on the GATB were almost identical. In fact, the GATB factor structures for the two groups were deemed equivalent. Lastly, numerous other studies have been documented with research originating in Austria, Brazil, Chile, Germany, India, Ireland, Italy, Japan, Korea, Netherlands, Singapore, Spain, and Switzerland (Droege, 1984). This growing international network of use can only serve to increase our confidence in the GATB's ability to measure aptitudes in many populations regardless of ethnicity.

New Computer Administered GATB

With the advent of efficient and relatively inexpensive computer technology, it seemed only a matter of time before the GATB would be transformed into a computer assessment format. The new GATB-Computerized Administration or GATB-CA was developed in 1995 (W. Martin, personal communication, January 31, 1995). The GATB-CA is contained on 17 diskettes and can be run on most desktop computers. Basically, it is an automated test administration program for subtests 1 to 7 of the GATB, Form B. It administers all or any of the subtests, computes raw

scores, reports number of errors and percentage correct by subtest, and converts raw scores to standard scores (Canadian Norms).

Administration is made easy for the assessor. All instructions are given by the computer via audio speakers or headphones. For individuals who are not familiar with computer use there is a front-end module that introduces the examinee to the computer and familiarizes the person on how to use the computer mouse. Following the instructions and introduction, the GATB-CA gives the standard practice exercises (same as conventional GATB), corrects any errors made during practice exercises, and then administers and times each subtest. When all of the selected subtests are completed, the scores are shown on the screen.

Although the introduction of the GATB-CA appears to be an advancement in GATB methodology, in the past there have been specific problems associated with changing GATB subtests. More specifically, when plastic pegboards (required for Parts 9 and 10 which are combined to form Aptitude M - Manual Dexterity) were produced in addition to the original wooden pegboards there was the question of whether this change would produce differences in subject scores. Two subsequent studies provided disturbing results. First, Kapes and Sievert (1973) compared the scores of 1,050 ninth grade high school students who completed the GATB using both plastic and wooden equipment. They found that the means taken together indicated a general mean difference of 5 points for Part 9 and 2.6 points for Part 10 ($p<.001$). Upon closer inspection of the data, one finds sizeable converted mean aptitude M differences ranging from a high of 26 converted score points to a low of 3. In all cases, the M aptitude differences favored the wooden equipment. Second, a restudy conducted by Trimmer and Klein (1974) found essentially the same results. Three hundred eighty-eight subjects (106 males, 282 females) were

randomly assigned to either plastic or wooden pegboard GATB groups. The final examination of the means for the entire group revealed an overall mean difference of 10.5 points (approximately 1 standard error of measurement) favoring the wooden apparatus which was significant at the .001 level. Thus, while the literature comparing plastic and wooden GATB pegboards is not always consistent, these studies indicate that even small alterations in the GATB's administrative format can result in major changes in obtained scores.

Another issue facing the new GATB-CA is the fact that when a conventional paper-and-pencil test is transferred to a computer for administration and scoring, there is no assurance that the scores achieved with the computer presentation will be comparable to those obtained with the conventional format (Greaud & Green, 1986). Previous research indicates that there can be significant differences in computer and conventional formats of the same test. To begin with, in early studies such as Wildgrube's (1982) comparison of computer modes with paper-and-pencil modes, it was found that there were no differences between modes with an arithmetic test, higher scores on the computerized version of a figural reasoning test, and higher scores with the paper-and-pencil version of a verbal test (cited in Lee, Moreno, & Sympson, 1986). Other researchers such as Lee, Moreno, and Sympson (1986) have found computer administrations to be less effective in assessments. In using both computer and conventional versions of the Experimental Arithmetic Reasoning Test (EXP-AR) to test 585 male subjects ages 18-25, they found that mode of test administration did have a statistically significant effect on test scores, with the mean score obtained by computer lower than that obtained by paper-and-pencil. Interestingly, when subjects were asked about individual test items it was reported that 21 of the 30 items were more difficult in the computer mode whereas

only 3 items were more difficult in the paper-and-pencil mode. Finally, Bunderson, Inouye, and Harvey (1989) completed a review of approximately 40 studies investigating the equivalence of computer-based versions to original versions of the same tests. They discovered that scores on computer-based versions were more often lower than higher relative to the conventional versions, although the differences were typically small and, in a few instances, computer-based versions even produced higher scores. Therefore, in looking at the GATB-CA it is vital to determine whether its use results in lower, higher, or equivalent subject scores.

A further issue that must be considered is the speed at which subjects complete the GATB-CA. In other computerized tests derived from conventional paper-and-pencil versions, it is often the case that subjects are able to respond more quickly, and in turn complete more items, when using the computer version. For example, one study (Lansman, Donaldson, Hunt, & Yantis, 1982) comparing performance on computerized and conventional versions of three tests (letter matching, sentence verification, and mental rotations) found that mean reaction time for correct answers on each of the computerized versions was faster than mean reaction time per correct item on the corresponding paper-and-pencil versions. A second study (Greaud & Green, 1986) found similar results. They contrasted two speeded clerical tests (numerical operations and coding speed) of the Armed Services Vocational Aptitude Battery (ASVAB) with their corresponding paper-and-pencil versions. Results showed that examinees were faster at pressing a button than at locating and marking a bubble on an answer sheet therefore allowing quicker responding on computer-presented clerical tests. Because time taken to respond to items is a critical component of examinees' scores it is an important factor that has to be addressed when evaluating the GATB-CA.

There are a number of possible reasons to explain changes in scores as a result of moving to computerized assessments. One obvious reason is that such a transformation may have important changes that affect the fundamental nature of the tasks involved. Another is that anxiety provoked by the computer ("computer phobia") and other affective reactions (positive or negative) that computers evoke in clients may also impact on test performance (McKee & Levinson, 1990). A third reason was reported by Bordieri and Musgrave (1989). They found that the older clients in their study perceived the computer exercises to be harder than the traditional academic testing style and more difficult to learn. Lastly, Lee et al. (1986) list a range of reasons in their investigation looking at the effects of mode of administration on test performance. Factors which led to differential performances between computer and paper-and-pencil tests included the amount of time available for testing, overall difficulty of the test, the cognitive processes required by the test, and the absence or presence of a human assessor. It is important to also remember that there may be complex interactions at play among these explanations. To summarize, there are a number of empirical questions that have to be answered before the GATB-CA becomes a viable and valuable alternative to the conventional GATB.

Despite the potentially negative consequences associated with moving towards computer administrations, there are numerous possible benefits that cannot be overlooked. The trend since the early 1980s is for researchers and counselors to take advantage of and adapt to this new technology. As there is an increase in the number of paper-and-pencil instruments that are being converted to computerized formats, more practitioners are realizing that this new format can save time and simplify administrative procedures (McKee & Levinson, 1990). A review by

Burkhead and Sampson (1985) of the computerized-testing literature summarized the advantages as follows: (a) flexibility in scheduling of testing; (b) rapid reporting of score, allowing immediate feedback and information for decision making; (c) efficiency and flexibility in manipulation of test data; (d) individualization of assessment; (e) cost effectiveness; and (f) reduced error rates.

Another positive finding is that the reliability and validity of computer assessments can equal or exceed that of conventional assessments. Greaud and Green (1986) in their comparison of speed tests found that reliability of scores on computer-presented clerical tests was at least as high as for paper-and-pencil tests. Moreover, the correlations between the two administrative modes were high. Another study (Reardon & Loughead, 1988) compared the paper-and-pencil Self-Directed Search (SDS) to the computerized SDS version and found no significant differences between the summary scale scores of subjects taking both versions. Therefore, these studies provide strong evidence to argue that conventional instruments can be converted to computerized assessments successfully without damaging reliability or validity of tests.

Finally, an important human element that should be noted is that there appears to be a consistent positive client response to the computerized format of tests (Bordieri & Musgrave, 1989; Chan et al., 1989; McKee & Levinson, 1990). Specifically, in a study intended to explore client perceptions, Bordieri and Musgrave (1989) found that clients reported significantly greater enjoyment with the computer tasks than the hardware (i.e., block assembly, wobbleboard assembly) tasks. Additionally, they also reported that the instructions for the computer tasks were easier to understand. A second example is presented by Reardon and Loughead (1988) who found that 86% of participants who took both a computerized and

paper-and-pencil version of the Self-Directed Search preferred the computerized method of administration.

In conclusion, in light of the fact that many common psychological and vocational tests are gradually switching to computer formats and the fact that there are real benefits to doing so, there is a definite need to produce a computerized GATB that is comparable to and as effective as its conventional form. The recent development of the GATB-CA was the beginning of this process but prior to this investigation there had been no research evaluating the equivalency of this instrument to the original GATB. Thus, one purpose of this study was to evaluate the equivalency of the GATB-CA in terms of subjects' aptitude scores and speed of responding (measured by the total number of items completed).

The GATB and the Prediction of Success in Academics and Vocations

One important area of research related to the GATB focuses on the instrument's ability to predict success in school or occupational settings. The GATB has been used extensively to predict academic success in various courses of study. Many of the researchers who have used the GATB for this purpose have found it to have moderate to good predictive ability. For instance, many studies have revealed that the GATB is correlated with general academic achievement. First, Hakstian and Bennet (1978) assessed 161 grade eleven students and showed that GATB Aptitude G scores moderately correlated (.48) with students' year-end grades. When data for the male students were examined alone, this correlation rose to .62. Second, Hanners and Bishop (1975) discovered that of the 9 aptitude scores measured by the GATB, the Aptitude G score was again the best predictor of success (.34) in eleventh and twelfth grade students (N = 172). They also noted that the best combination of aptitudes for predicting success in classes was the G, N, P, F, and M

aptitude scores (.38). A third study (Moore & Davies, 1984) which investigated factors relating to General Educational Development (GED) scores in 224 students entering GED preparatory classes, found that the Aptitude G score of the GATB was significantly correlated (.56) with the overall GED scores. Thus, it can be seen that the GATB scores (especially G scores) can be used quite efficiently to predict success in academic achievement.

In addition, the GATB has been shown to be useful in predicting success among students who are undergoing training associated with a particular occupation. For example, Weber, King, and Pitts (1973) conducted a study to determine variables associated with success in schools of practical nursing. Data from 922 subjects dispersed among 25 nursing schools were analyzed. They discovered that of the 4 GATB subtests used as variables (G, V, Q, and K) two were moderately and significantly ($p<.01$) correlated with the criterion, State Board Examination scores for nursing. These were GATB Aptitude G (.35) and Aptitude V (.45) scores. Another example comes from a study (Dolke & Sharma, 1975) using students randomly selected from higher level School of Architecture classes. Here it was found that many GATB aptitude scores were significantly correlated with the criterion variable of average overall final grade: G(.55), V(.34), N(.43), S(.39), P(.25), and Q(.37). These two studies clearly indicate that the predictive ability of the GATB can help serve as a device for better selection of students into a specialized program of study. The second purpose of the present study then was to evaluate and compare both GATB versions abilities to predict success in a population of subjects all enrolled in the same course of study.

To summarize, there were two major aims of the present study. One was to evaluate the equivalency of the new GATB-Computerized Administration to the

conventional GATB with respect to subjects scores and speed of responding (measured by the total number of items completed). The second was to investigate whether the scores from either of these GATB versions were predictive of subjects' success in a specific course of study.

Method

Subjects

Subjects (N = 62: 14 males and 48 females; age: M = 20.86 years, SD = 3.13 years) were recruited from students enrolled in undergraduate introductory psychology courses. Subjects were randomly divided into two groups, one of which completed the computer administered GATB followed by the paper-and-pencil GATB (n = 30: 6 males and 24 females; age: M = 20.25 years, SD = 2.26 years) while the other completed the paper-and-pencil GATB followed by the computer administered GATB (n = 32: 8 males and 24 females; age: M = 21.45 years, SD = 3.74 years).

Informed consent (see Appendix A) and voluntary participation was obtained prior to subjects participating in the study. All subjects received two credits toward their final course grade for participation.

Materials and Apparatus

General Aptitude Test Battery (GATB) Form B. Only the paper-and-pencil subtests, which consists of Parts 1 to 7 and comprise the 6 Aptitudes of Intelligence, Verbal Aptitude, Numerical Aptitude, Spatial Aptitude, Form Perception, and Clerical Perception, were used in this study.

General Aptitude Test Battery - Computerized Administration (GATB-CA). This program was a computerized translation of the conventional GATB, Form B, Parts 1 to 7. It comprised the same 6 GATB Aptitudes and automatically computed raw scores and standard scores.

Computer System. The configuration of the computer system used to run the GATB-CA program was as follows:

- 486 DX4 90 CPU
- 1 MB Video Card
- 15" Compaq Presario 1500 Colour Monitor
- 16 MB RAM
- Hard Drive with 10 millisecond access
- Windows Sound System 16-bit audio card
- External speakers
- Compaq Desktop Mouse

Comparable computer systems can be used to run the GATB-CA program but two requirements must be met: GATB-CA must run in at least 800 X 600 resolution mode and must also be presented on at least a 15-inch monitor.

Design and Procedures

The first part of the study, which focused on determining the equivalence of the two GATB test versions, involved a random assignment counterbalanced design. All subjects were assessed with both versions of the GATB, but in order to control for order effects half were first administered the conventional GATB and the other half were first administered the GATB-CA. In addition, to minimize any practice effects there was a minimum time period of 3 weeks between initial testing and retesting.

The conventional GATB was administered in small group sessions (4 to 8 individuals per group) with strict adherence to the procedures outlined in the GATB manual maintained (Manual for the GATB, Section I, 1986). Moreover, all group sessions were conducted and supervised by the same test administrator. In

contrast, the computerized version of the GATB (GATB-CA) was completed by subjects individually with the GATB-CA presenting standardized instructions based closely on the conventional GATB procedures.

For the second part of the study, subjects' final introductory psychology course grades were obtained from course instructors. These grades were then used to determine if any relationship existed between GATB aptitude scores and academic success in the specific course of study.

At the end of the study, subjects were debriefed (see Appendix B) and offered any information regarding the overall results of the completed research.

Results

All data entry and statistical analyses were performed using SPSS for Windows Standard Version 6.1.3. Prior to analysis, the data file was examined and deemed free of any data entry errors and critical missing values.

Preliminary Analyses

During the initial screening of the data, four univariate outliers were detected. To reduce their influence on the results, subject scores were altered to remain within 3 z scores of the variable mean as suggested by Tabachnick and Fidell (1989).

Due to the counterbalanced design of the study, subjects were tested in two groupings. Group 1 completed the computerized GATB followed by the conventional GATB and Group 2 completed the tests in the reverse order. There were no statistically significant group differences with respect to mean age, $t(46) = -1.47$, ns, and mean psychology final course grade, $t(49) = 1.40$, ns. The only significant group difference was in the mean interval between testings in days of the two groups. Group 1 ($M = 29.10$, $SD = 3.51$) had fewer days between testings than Group 2 ($M = 55.88$, $SD = 5.77$), $t(52) = -22.24$, $p < .001$.

Main Analyses

Before proceeding to present the main findings from the data, an important issue must be addressed concerning multiple comparison procedures. Due to the relatively simple design of this study, the entire analysis of data involved using a large number of bivariate correlations and t-tests. As a result, there was a need to adjust the criterion for significance to account for the large number of comparisons performed. In accordance with the Bonferroni test, the appropriate per comparison significance level for this study should be .0003. Using this significance level would reduce the chance of making a Type I error but more importantly, would increase the probability of not detecting any "real" difference that may exist. In this study, the primary focus is to determine whether there is any significant difference between the conventional GATB and the newly developed computerized GATB. In most research, the desired result is to find significant differences and reject the null hypothesis. By contrast, in the present study, the desired result would be to not find any significant differences between the two versions of the GATB and conclude that the two tests are equivalent. Therefore, by using the conservative significance level set by the Bonferroni test (.0003), there would be an increased probability of making the more critical error of not detecting any significant differences between the GATB and the GATB-CA when one genuinely exists and erroneously concluding that the two tests are equivalent. Considering the circumstances, it was decided that the best solution to this dilemma would be to compromise and use a significance level between .05 and .0003. The criterion for significance per comparison in this study was set at .01.

Comparison 1: Computer vs. Conventional GATB

The first analysis performed compared the GATB-CA to the GATB with respect to the 20 target variables of interest: raw score on the seven subtests, number of items completed on the seven subtests, and score on the six aptitudes (G, N, P, Q, S, V). This comparison consisted of paired samples t-tests using all subjects ($N = 62$). Results showed that subjects consistently performed better on many of the GATB-CA indices. Specifically, subjects scored significantly higher on four of the seven subtests, completed more items on five of the seven subtests, and scored higher on four of the seven aptitudes as shown in Table 2.

Comparison 2: Computer vs. Conventional GATB

This comparison used a between-subjects design to compare the scores of subjects completing the GATB and GATB-CA for the first time. Independent samples t-tests were conducted to determine if there were any significant differences between Group 1 scores on the GATB-CA ($n = 30$) and Group 2 scores on the conventional GATB ($n = 32$). Similar to Comparison 1, subjects who completed the GATB-CA, showed consistently higher performances on eight of the 20 variables. Specifically, subjects demonstrated higher scores on two of the seven subtests, completed more items on three of the seven subtests, and received higher scores on three of the six aptitudes as shown in Table 3.

Comparison 3: Computer vs. Conventional GATB

A between-subjects design was again used to compare the scores of subjects completing the GATB and GATB-CA after they had previously taken one version of the test. Independent samples t-tests were used to determine if any significant

Table 2

Within-Subjects Comparison Between GATB-CA and Conventional GATB on Subtest Scores, Subtest Items Completed, and Aptitude Scores (N = 62)

Target Variable	Test Means		t-value
	GATB-CA	Conventional GATB	
Subtest 1 Score	79.16	56.23	11.89**
Subtest 2 Score	23.15	21.29	4.82**
Subtest 3 Score	23.68	22.47	1.85
Subtest 4 Score	25.40	24.02	3.05*
Subtest 5 Score	34.92	32.77	3.21*
Subtest 6 Score	10.98	10.53	1.75
Subtest 7 Score	32.34	33.16	-1.29
No. Items Completed in Subtest 1	84.35	59.29	12.23**
No. Items Completed in Subtest 2	25.58	23.50	6.01**
No. Items Completed in Subtest 3	28.66	26.69	2.78*
No. Items Completed in Subtest 4	34.81	30.45	5.82**
No. Items Completed in Subtest 5	35.92	33.60	3.24*
No. Items Completed in Subtest 6	13.29	12.73	1.97
No. Items Completed in Subtest 7	35.48	35.03	.78
Aptitude G Score	99.84	96.73	2.86*
Aptitude N Score	97.45	91.31	4.79**
Aptitude P Score	111.00	108.97	1.28
Aptitude Q Score	139.18	104.35	11.88**
Aptitude S Score	110.68	106.85	1.91
Aptitude V Score	99.34	95.92	3.11*

*$p < .01$.

**$p < .001$.

Table 3

First Between-Subjects Comparison Between GATB-CA and Conventional GATB on Subtest Scores, Subtest Items Completed, and Aptitude Scores

Target Variable	Test Means: Group 1 GATB-CA (n=30)	Test Means: Group 2 Conventional GATB (n=32)	t-value
Subtest 1 Score	76.03	53.38	6.85**
Subtest 2 Score	22.80	20.22	2.51
Subtest 3 Score	21.93	20.00	1.50
Subtest 4 Score	26.00	21.69	3.15*
Subtest 5 Score	33.30	31.72	1.01
Subtest 6 Score	11.00	9.78	2.13
Subtest 7 Score	31.33	32.53	-.74
No. Items Completed in Subtest 1	81.37	56.81	6.62**
No. Items Completed in Subtest 2	25.30	22.75	3.09*
No. Items Completed in Subtest 3	27.20	24.19	2.44
No. Items Completed in Subtest 4	34.43	27.88	5.03**
No. Items Completed in Subtest 5	34.23	32.38	1.16
No. Items Completed in Subtest 6	12.83	12.34	.87
No. Items Completed in Subtest 7	34.83	34.69	.10
Aptitude G Score	99.80	90.94	3.41*
Aptitude N Score	96.40	87.22	2.52
Aptitude P Score	107.30	106.63	.18
Aptitude Q Score	134.37	100.06	6.83**
Aptitude S Score	105.33	99.28	1.53
Aptitude V Score	100.80	90.34	3.17*

*$p < .01$.

**$p < .001$.

differences existed between Group 1 scores on the conventional GATB (n=30) and Group 2 scores on the GATB-CA (n=32). The results showed that subject scores were equivalent for both versions of the GATB with the exception of one subtest. Subjects performed significantly better on subtest 1 as demonstrated by higher raw scores, more items completed, and higher scores on aptitude Q which corresponds to subtest 1 (see Table 4).

Comparison 4: Computer vs. Conventional GATB

In order to assess whether the order in which subjects completed the two versions of the GATB affected their performance on the GATB and GATB-CA, two separate but related comparisons were performed.

First, paired samples t-tests were used to detect any significant differences between scores on the GATB-CA and conventional GATB of Group 1 subjects (who completed the GATB-CA followed by the GATB). Results indicated that there were significant differences that led subjects to perform better on certain parts of the computerized GATB while performing better on certain other parts of the conventional GATB. Specifically, Group 1 subjects scored higher and completed more items on the GATB-CA subtest 1 which resulted in higher scores on its corresponding aptitude Q. Opposite to this trend, subjects scored higher on subtests 3 and 7 of the conventional GATB and received higher scores on subtest 3's corresponding aptitude S as shown in Table 5.

The second comparison involved paired samples t-tests to detect significant differences between scores on the GATB-CA and conventional GATB of Group 2 subjects (who completed the GATB followed by the GATB-CA). The results of this analysis revealed that Group 2 subjects performed significantly better on almost all indices of the GATB-CA (18 of the 20 target variables). Specifically, Group 2

Table 4

Second Between-Subjects Comparison Between GATB-CA and Conventional GATB on Subtest Scores, Subtest Items Completed, and Aptitude Scores

Target Variable	Test Means: Group 1 Conventional GATB (n=30)	Test Means: Group 2 GATB-CA (n=32)	t-value
Subtest 1 Score	59.27	82.09	-7.15**
Subtest 2 Score	22.43	23.47	-1.00
Subtest 3 Score	25.10	25.31	-.14
Subtest 4 Score	26.50	24.84	1.12
Subtest 5 Score	33.90	36.44	-1.51
Subtest 6 Score	11.33	10.97	.64
Subtest 7 Score	33.83	33.28	.32
No. Items Completed in Subtest 1	61.93	87.16	-7.18**
No. Items Completed in Subtest 2	24.30	25.84	-1.79
No. Items Completed in Subtest 3	29.37	30.03	-.47
No. Items Completed in Subtest 4	33.20	35.16	-1.33
No. Items Completed in Subtest 5	34.90	37.50	-1.43
No. Items Completed in Subtest 6	13.13	13.72	-.91
No. Items Completed in Subtest 7	35.40	36.09	-.41
Aptitude G Score	102.90	99.88	1.05
Aptitude N Score	95.67	98.44	-.77
Aptitude P Score	111.47	114.47	-.70
Aptitude Q Score	108.93	143.69	-7.18**
Aptitude S Score	114.93	115.69	-.16
Aptitude V Score	101.87	97.97	1.10

*$p < .01$.

**$p < .001$.

Table 5

Group 1 Comparison Between GATB-CA and Conventional GATB on Subtest Scores, Subtest Items Completed, and Aptitude Scores (n=30)

Target Variable	Test Means: Group 1 GATB-CA	Test Means: Group 1 Conventional GATB	t-value
Subtest 1 Score	76.03	59.27	5.88**
Subtest 2 Score	22.80	22.43	.92
Subtest 3 Score	21.93	25.10	-5.86**
Subtest 4 Score	26.00	26.50	-.88
Subtest 5 Score	33.30	33.90	-.69
Subtest 6 Score	11.00	11.33	-1.08
Subtest 7 Score	31.33	33.83	-2.91*
No. Items Completed in Subtest 1	81.37	61.93	6.34**
No. Items Completed in Subtest 2	25.30	24.30	1.91
No. Items Completed in Subtest 3	27.20	29.37	-2.59
No. Items Completed in Subtest 4	34.43	33.20	1.37
No. Items Completed in Subtest 5	34.23	34.90	-.72
No. Items Completed in Subtest 6	12.83	13.13	-.90
No. Items Completed in Subtest 7	34.83	35.40	-.72
Aptitude G Score	99.80	102.90	-2.61
Aptitude N Score	96.40	95.67	.65
Aptitude P Score	107.30	111.47	-2.03
Aptitude Q Score	134.37	108.93	5.88**
Aptitude S Score	105.33	114.93	-5.81**
Aptitude V Score	100.80	101.87	-.77

*$p < .01$.

**$p < .001$.

subjects scored higher on six of the seven subtests, completed more items on six of the seven subtests, and received higher scores on all of the aptitudes as shown in Table 6.

Comparison 5: Computer vs. Computer GATB

This comparison, which involved using independent samples t-tests, contrasted the GATB-CA scores of Group 1 (no previous exposure to the GATB) with Group 2 (previously completed the conventional GATB) to test for any significant practice effects in Group 2 GATB-CA scores. The results showed that although Group 2 subjects had higher mean scores on 17 of the 20 target variables, these increases were not statistically significant (see Table 7).

Comparison 6: Conventional vs. Conventional GATB

Similar to the last comparison, this comparison again used independent samples t-tests to contrast the conventional GATB scores of Group 1 (previously completed the computerized GATB) with Group 2 (no previous exposure to the GATB) to test for any significant practice effects in Group 1 conventional GATB scores. In this analysis, Group 1 subjects had higher mean scores on all 20 of the target variables. Upon closer inspection of the results, it was found that Group 1 performed significantly better than Group 2 on two subtests of the conventional GATB and the three aptitudes that are influenced by these two subtests. Specifically, Group 1 subjects scored higher on and completed more items in subtests 3 and 4 and as a result also received higher scores in the three related aptitudes of G, S, and V, suggesting a strong practice effect from previously being exposed to the GATB-CA (see Table 8).

Table 6

Group 2 Comparison Between GATB-CA and Conventional GATB on Subtest Scores, Subtest Items Completed, and Aptitude Scores (n=32)

	Test Means		
Target Variable	Group 2 GATB-CA	Group 2 Conventional GATB	t-value
Subtest 1 Score	82.09	53.38	13.06**
Subtest 2 Score	23.47	20.22	5.97**
Subtest 3 Score	25.31	20.00	10.56**
Subtest 4 Score	24.84	21.69	5.80**
Subtest 5 Score	36.44	31.72	6.11**
Subtest 6 Score	10.97	9.78	3.24*
Subtest 7 Score	33.28	32.53	.87
No. Items Completed in Subtest 1	87.16	56.81	12.46**
No. Items Completed in Subtest 2	25.84	22.75	8.04**
No. Items Completed in Subtest 3	30.03	24.19	10.73**
No. Items Completed in Subtest 4	35.16	27.88	7.88**
No. Items Completed in Subtest 5	37.50	32.38	6.22**
No. Items Completed in Subtest 6	13.72	12.34	3.33*
No. Items Completed in Subtest 7	36.09	34.69	1.72
Aptitude G Score	99.88	90.94	8.89**
Aptitude N Score	98.44	87.22	6.07**
Aptitude P Score	114.47	106.63	4.12**
Aptitude Q Score	143.69	100.06	13.05**
Aptitude S Score	115.69	99.28	10.70**
Aptitude V Score	97.97	90.34	5.78**

*$p < .01$.

**$p < .001$.

Table 7

Comparison Between Group 1 GATB-CA and Group 2 GATB-CA Subtest Scores, Subtest Items Completed, and Aptitude Scores

Target Variable	Test Means: Group 1 GATB-CA (n=30)	Test Means: Group 2 GATB-CA (n=32)	t-value
Subtest 1 Score	76.03	82.09	-1.68
Subtest 2 Score	22.80	23.47	-.60
Subtest 3 Score	21.93	25.31	-2.35
Subtest 4 Score	26.00	24.84	.78
Subtest 5 Score	33.30	36.44	-1.97
Subtest 6 Score	11.00	10.97	.06
Subtest 7 Score	31.33	33.28	-1.17
No. Items Completed in Subtest 1	81.37	87.16	-1.44
No. Items Completed in Subtest 2	25.30	25.84	-.61
No. Items Completed in Subtest 3	27.20	30.03	-1.95
No. Items Completed in Subtest 4	34.43	35.16	-.49
No. Items Completed in Subtest 5	34.23	37.50	-1.94
No. Items Completed in Subtest 6	12.83	13.72	-1.41
No. Items Completed in Subtest 7	34.83	36.09	-.79
Aptitude G Score	99.80	99.88	-.03
Aptitude N Score	96.40	98.44	-.54
Aptitude P Score	107.30	114.47	-1.83
Aptitude Q Score	134.37	143.69	-1.70
Aptitude S Score	105.33	115.69	-2.35
Aptitude V Score	100.80	97.97	.79

*$p < .01$.

**$p < .001$.

Table 8

Comparison Between Group 1 Conventional GATB and Group 2 Conventional GATB Subtest Scores, Subtest Items Completed, and Aptitude Scores

Target Variable	Test Means: Group 1 Conventional GATB (n=30)	Test Means: Group 2 Conventional GATB (n=32)	t-value
Subtest 1 Score	59.27	53.38	2.12
Subtest 2 Score	22.43	20.22	2.32
Subtest 3 Score	25.10	20.00	3.75**
Subtest 4 Score	26.50	21.69	3.54*
Subtest 5 Score	33.90	31.72	1.32
Subtest 6 Score	11.33	9.78	2.59
Subtest 7 Score	33.83	32.53	.79
No. Items Completed in Subtest 1	61.93	56.81	1.67
No. Items Completed in Subtest 2	24.30	22.75	1.93
No. Items Completed in Subtest 3	29.37	24.19	4.41**
No. Items Completed in Subtest 4	33.20	27.88	4.14**
No. Items Completed in Subtest 5	34.90	32.38	1.45
No. Items Completed in Subtest 6	13.13	12.34	1.34
No. Items Completed in Subtest 7	35.40	34.69	.44
Aptitude G Score	102.90	90.94	4.48**
Aptitude N Score	95.67	87.22	2.46
Aptitude P Score	111.47	106.63	1.15
Aptitude Q Score	108.93	100.06	2.10
Aptitude S Score	114.93	99.28	3.74**
Aptitude V Score	101.87	90.34	3.53*

*$p < .01$.

**$p < .001$.

Relationship between GATB Aptitude Scores and Academic Success

Pearson correlation coefficients were computed to determine relationships between GATB aptitude scores and academic success in an undergraduate level course. Results showed that two out of the six aptitudes of both the GATB and GATB-CA correlated highly with final course grades of subjects. These were Intelligence (G) and Verbal Aptitude (V) as shown in Table 9.

Supplementary Analyses

In addition to the main analyses, some additional analyses were conducted in order to better explain the results and expand on the equivalence of the two versions of the GATB.

First, Pearson correlation coefficients were computed to determine if there were any relationships between Group 1 interval between testings and 1) subjects scores on the second GATB test completed (conventional GATB) and 2) subjects change in scores from their initial GATB assessment to their retesting on the GATB. Findings revealed that there were no significant correlations on any of the 20 target variables. Similarly, correlations between Group 2 interval between testings and 1) subjects scores on the second GATB test completed (computerized GATB) and 2) subjects change in scores from their initial GATB assessment to their retesting on the GATB again showed no significant correlations on any of the 20 target variables.

Second, correlations were computed to determine if any relationship existed between subject scores of the 20 target variables of the first GATB testing and subject scores of the same 20 target variables of the GATB retesting. Separate analyses were performed for each of the two groupings of subjects. In Group 1, all variables were significantly correlated ($p<.001$) with the corresponding variable in the second testing with the exception of the score and number of items completed

Table 9

Correlation Coefficients Between GATB Aptitude Scores and Academic Success in a University Course (N = 62)

Target Variable	r between GATB-CA and Final Grade	r between Conventional GATB and Final Grade
Aptitude G (Intelligence)	.41*	.48**
Aptitude N (Numerical Aptitude)	.18	.26
Aptitude P (Form Perception)	.06	.10
Aptitude Q (Clerical Perception)	.20	.17
Aptitude S (Spatial Aptitude)	-.01	.05
Aptitude V (Verbal Aptitude)	.48**	.52**

*$p < .01$.

**$p < .001$.

on subtest 1 and the score of the corresponding GATB aptitude Q which were not significant. In Group 2 these results were replicated, with all target variables from the first GATB testing being significantly correlated ($p < .01$) with their corresponding variable in the GATB retesting. Again, the exception to this finding was the correlations relating the score and number of items completed on subtest 1 and the score of the corresponding GATB aptitude Q which were not significant.

Third, the coefficients of stability using parallel forms (GATB and GATB-CA) were computed for the six target GATB aptitudes. These values were then compared to the coefficients of stability using the same form (conventional GATB at two separate testings) for the six target GATB aptitudes as provided by the Manual for the GATB, Section III, 1980, p. 258. Both analyses yielded similar values as shown in Table 10.

Finally, to compare the underlying factor structures of the two GATB versions, principal components analyses with varimax rotations and using the Eigenvalue of 1 criterion were performed on subtests from the GATB-CA and the original GATB.

With respect to the GATB-CA, this analysis yielded a two-factor solution. Factor 1, which accounted for 42.4% of the total variance, loaded on the subtests that measure visual-spatial skills (subtests 1, 3, 5, 7). Factor 2, which accounted for 21.5% of the variance, loaded on the subtests that measure general intelligence (subtests 2, 6). The factor loadings are presented in Table 11.

A similar underlying structure can be seen in the conventional GATB as the factor analysis again extracted a two-factor solution. Factor 1, with loadings on subtests 1, 3, 5, and 7, accounted for 44.3% of the total variance, and Factor 2, with loadings on subtests 2, 4, and 6, accounted for 22.9% of the variance (see Table 11).

Table 10

Coefficients of Stability for GATB Aptitudes using Parallel Forms (GATB-CA and Conventional GATB) and the Same Form (Conventional GATB)

Target Variable	*r* using Parallel Forms[a] (N = 62)	*r* using the Same Form[b] (N = 231)
Aptitude G (Intelligence)	.73**	.87**
Aptitude N (Numerical Aptitude)	.76**	.85**
Aptitude P (Form Perception)	.70**	.72**
Aptitude Q (Clerical Perception)	.32	.77**
Aptitude S (Spatial Aptitude)	.62**	.87**
Aptitude V (Verbal Aptitude)	.81**	.89**

Note. *$p < .01$. **$p < .001$.

[a] $N = 62$. Mean interval between testings = 43 days.

[b] $N = 231$. Mean interval between testings = 90 days. From the Manual for the GATB, Section III, 1980, p. 258.

Table 11

Factor Loadings on the Varimax Rotated Factor Matrix for the GATB-CA and Conventional GATB Subtests

	Factor 1	Factor 2
GATB-CA		
Subtest 1	.71	.33
Subtest 2	-.01	.88
Subtest 3	.77	.07
Subtest 4	.39	.35
Subtest 5	.83	-.15
Subtest 6	.17	.87
Subtest 7	.82	.22
Conventional GATB		
Subtest 1	.78	.10
Subtest 2	.07	.87
Subtest 3	.66	.30
Subtest 4	.38	.53
Subtest 5	.89	-.07
Subtest 6	.03	.90
Subtest 7	.85	.18

Discussion

The present study was designed to evaluate whether the newly developed GATB-CA could be used as an acceptable alternative to the original GATB. Evaluation of the equivalency of the two GATB versions focused on three specific areas: subtest and aptitude scores, speed of the completion of subtest items, and the power to predict academic success.

Subtest and Aptitude Scores

The two analyses that best answer the question of whether the GATB-CA and conventional GATB produce equivalent test scores are the within-subjects comparison using all 62 subjects (see Table 2) and the between-subjects comparison where subjects were introduced to the GATB test for the first time (see Table 3).

By combining the results of both of these analyses, it was found that the GATB-CA was equivalent to the original GATB on the scores of subtests 3, 6, and 7, and the scores of aptitudes P and S. For the remaining scores, subjects significantly performed better on the GATB-CA in either one or both of the analyses. The subtests and aptitudes that were consistently significantly different in favour of the GATB-CA were subtests 1 and 4 and aptitudes G, Q, and V. Therefore, the conclusion is that the GATB-CA cannot be used as an acceptable alternative to the original GATB without some revisions to correct for these differences. The findings that subjects scored higher on the computerized version of the GATB is contrary to the results of the Bunderson et al. (1989) review investigating the equivalence of computer-based versions to original versions of the same tests. They found that subjects generally scored lower on computer-based tests relative to their conventional counterparts.

Speed of the Completion of Subtest Items

To evaluate whether the GATB-CA was equivalent to the original GATB with respect to the speed in which subjects completed subtest items, the same two analyses were used: the within-subjects comparison using all 62 subjects (see Table 2) and the between-subjects comparison where subjects were introduced to the GATB test for the first time (see Table 3).

Again, by combining the results of both analyses it was found that the GATB-CA was equivalent to the original GATB in terms of item completion speed on subtests 6 and 7. However, for the remaining five subtests, subjects completed significantly more subtest items during the GATB-CA in either one or both analyses. Subtests in which subjects consistently completed more items in the GATB-CA version were subtests 1, 2, and 4.

The finding that subjects responded quicker to subtest items when using a computerized version of a test was consistent with previous findings in the literature. For instance, one study (Lansman et al., 1982) showed that mean reaction times for correct answers on computerized versions of tests were faster than mean reaction times per correct item on the corresponding paper-and-pencil tests. A second study (Greaud & Green, 1986) found similar results with subjects responding more quickly on computer-presented clerical tests compared to their original paper-and-pencil versions.

There are a number of possible reasons that may account for this difference in speed of item completion. The first reason concerns the obvious mechanical difference in how one responds to test items. Greaud and Green (1986) theorized on the basis of their study that it is simply faster to click a computer mouse button than it is to locate and then bubble in a space on an answer sheet. Additionally, the

computer format removes the need to constantly alternate attention from one location (test booklet) to another (answer sheet) which should result in a reduction in the time needed to respond to items. A third explanation is that in today's rapidly advancing world most people, especially students, are becoming very comfortable and proficient in using computers. Many times during the present study subjects stated having preferred the computerized GATB over the conventional GATB after having completed both versions. These subjects mentioned that the GATB-CA directions were easier to follow and incorrect responses were quicker to correct using the mouse. Finally, another factor that may have influenced item response speed is the fact that in the GATB-CA the timer that determines how much time is left for each subtest is continuously displayed at the bottom of the screen. This may elevate the number of items subjects complete if subjects randomly guess responses as the timer nears the end.

The Power to Predict Academic Success

In order to test whether either of the GATB versions was able to predict success in a university level course of study, correlation coefficients were computed between subjects' final grade and the six aptitude scores of both the GATB-CA and the conventional GATB (see Table 9). The results showed that both GATB versions were equivalent in their abilities to predict academic success. Specifically, aptitudes G and V in both GATB versions moderately and significantly correlated with final course grades.

These findings strongly resemble those of past research investigating GATB aptitudes and academic success (Dolke & Sharma, 1975; Hakstian & Bennet, 1978; Hanners & Bishop, 1975; Moore & Davies, 1984; Weber et al., 1973). In all of these studies, GATB aptitudes G and V were found to be good, if not the best,

predictors of subjects' academic success with G correlations ranging from .34 to .56 and V correlations ranging from .34 to .45.

Additional Findings

Some of the supplementary analyses performed lend support to the notion that the GATB-CA may be closely related to but not equivalent to the original GATB. For instance, subjects who performed well on the GATB-CA also did well on the conventional GATB relative to others. This was indicated by the analysis correlating the 20 target variables of the first GATB testing with the same 20 variables of the second GATB testing for both Group 1 and 2 separately. In all cases, with the exception of subtest 1 and its corresponding aptitude Q, subjects' first scores were significantly correlated to their second scores.

In another analysis, the GATB-CA appeared to serve as an adequate substitute for the original GATB when calculating coefficients of stability from initial testing to retesting (see Table 10). For all of the target aptitudes with the exception of aptitude Q, the GATB-CA was able to match the level of stability of scores from initial testing to retesting.

Finally, in the factor analyses conducted to compare the underlying structures of the GATB-CA and conventional GATB, it was revealed that both versions had essentially the same factor structure (see Table 11). Subtests 1 to 7 of both GATB versions can be reduced to two factors, one which can be labelled Visual-Spatial Skills (Factor 1) and the other General Intelligence (Factor 2). This similarity in structure provides additional support to show that the two GATB versions are closely related.

Limitations and Unexpected Findings

A shortcoming in the present study was that there was a significant difference in the interval between testings of Group 1 and Group 2 subjects. This was due to a short period of minor hardware and software problems running the GATB-CA. To assess whether this difference significantly influenced subjects' scores two analyses were performed. Correlation coefficients were computed between the interval in days and 1) subjects scores on the second GATB test completed and 2) subjects change in scores from their initial GATB assessment to their retesting on the GATB for both Groups 1 and 2 separately. Because there were no significant findings on any of the 20 target variables for either group, it was concluded that this difference in interval between testings did not play a critical role in influencing subjects' scores.

Another shortcoming is the simplicity of the analyses performed in this study. Because this study was exploratory in nature and aimed at providing a basic initial comparison between the GATB-CA and original GATB, few in-depth psychometric techniques were employed. Future studies which seek to evaluate revised versions of the GATB-CA should conduct more advanced and in-depth psychometric analyses (e.g., item analysis, testing of the psychometric properties of individual subtests).

Although not specifically a shortcoming, another finding should be addressed. It appeared from the analyses that there was a differential practice effect between completing the GATB-CA and completing the conventional GATB. This can be deduced from the results of two analyses. First, when the GATB-CA scores of Group 1 (no previous exposure to GATB) are compared to the GATB-CA scores of Group 2 (previously completed conventional GATB) there are no significant differences (see Table 7). This means that any practice effect that occurred from previously taking the conventional GATB must in effect be small. Second, when the

conventional GATB scores of Group 1 (previously completed GATB-CA) are compared to the conventional GATB scores of Group 2 (no previous exposure to GATB) there are a number of significant differences favouring Group 1 scores (see Table 8). The resulting interpretation is that the practice effects that occurred from previous exposure to the GATB-CA must in effect be quite large.

Taking these differential practice effects into account made the interpretation of other analyses substantially simpler. For example, the relatively small number of significant differences between Group 1 GATB-CA scores and their conventional GATB scores (see Table 5) was probably due to the large practice effect from first taking the GATB-CA carrying over to considerably raise the level of their performance on the more difficult conventional GATB therefore making the two tests comparable. Conversely, the large number of significant differences between Group 2 GATB-CA scores and their conventional GATB scores (see Table 6) could likely be attributed to the small practice effect gained from first taking the conventional GATB adding to their performance on the relatively easier GATB-CA to exacerbate any differences that may already exist between the two GATB versions. Lastly, the high level of equivalence in the between-subjects comparison of Group 1 conventional GATB scores and Group 2 GATB-CA scores (see Table 4) can again be explained by the large practice effect experienced by Group 1 subjects carrying over to make their conventional GATB scores comparable to Group 2 subjects' GATB-CA scores (which are relatively unaffected by the small practice effect gained from their previous exposure to the conventional GATB).

Conclusion

The analyses in this study showed that although the GATB-CA closely resembled the conventional GATB in many areas, it was not deemed to be equivalent to the original as a whole. The parts of the GATB-CA that could be considered equal to the original GATB in terms of scoring and item completion are subtests 6 and 7 and aptitudes P and S. In contrast, the parts of the GATB-CA that are consistently and significantly different from the original in terms of scoring and item completion are subtests 1 and 4 and aptitudes G, Q, and V. It was found that subtest 1 and its corresponding aptitude Q showed the most deviation from the original GATB with subjects performing markedly better on the GATB-CA version of this subtest.

Surprisingly, despite its differences from the original GATB, the GATB-CA was still able to predict academic success of subjects with nearly the same level of confidence as the original GATB. This reinforced the conclusion that although the GATB-CA needs some revision and adjustment in order to become an acceptable alternative to the conventional GATB, this initial release was not too far off the mark.

References

Baydoun, R. B., & Neuman, G. A. (1992). The future of the General Aptitude Test Battery (GATB) for use in public and private testing. Journal of Business and Psychology, 7(1), 81-91.

Bemis, S. E. (1968). Occupational validity of the General Aptitude Test Battery. Journal of Applied Psychology, 52(3), 240-244.

Bordieri, J. E., & Musgrave, J. (1989). Client perceptions of the Microcomputer Evaluation and Screening Assessment. Rehabilitation Counseling Bulletin, 32(4), 342-345.

Bunderson, C. V., Inouye, D. K., & Harvey, J. B. (1989). The four generations of computerized educational measurement. In R. L. Linn (Ed.), Educational measurement (3rd ed., pp. 367-408). New York: Macmillan.

Burkhead, E. J., & Sampson, J. P. (1985). Computer-assisted assessment in support of the rehabilitation process. Rehabilitation Counseling Bulletin, 28, 262-274.

Chan, F., Lam, C. S., Leahy, M. J., Parker, H. J., & Wong, D. W. (1989). Computer applications in vocational evaluation: Current development and future directions. Vocational Evaluation and Work Adjustment Bulletin, 22(3), 109-115.

Dagenais, F. (1990). General Aptitude Test Battery factor structures for Saudi Arabian and American samples: A comparison. Psychology and Developing Societies, 2(2), 217-240.

Dolke, A. M., & Sharma, R. S. (1975). General Aptitude Test Battery (GATB) as a predictor of academic success in architectural courses. Indian Journal of Psychology, 50(2), 163-171.

Droege, R. C. (1984). The General Aptitude Test Battery and its international use. International Review of Applied Psychology, 33(3), 413-416.

Greaud, V. A., & Green, B. F. (1986). Equivalence of conventional and computer presentation of speed tests. Applied Psychological Measurement, 10(1), 23-34.

Hakstian, A. R., & Bennet, R. W. (1978). Validity studies using the Comprehensive Ability Battery (CAB): II. Relationships with the DAT and GATB. Educational and Psychological Measurement, 38(4), 1003-1015.

Hanners, E. F., & Bishop, T. D. (1975). Predicting success in vocational training by using the GATB. Journal of Employment Counseling, 12(2), 59-65.

Howe, M. A. (1975). General Aptitude Test Battery: An Australian empirical study. American Psychologist, 10(1), 32-44.

Janikowski, T. P., Berven, N. L., & Bordieri, J. E. (1991). Validity of the Microcomputer Evaluation Screening and Assessment aptitude scores. Rehabilitation Counseling Bulletin, 35(1), 38-51.

Kapes, J. T., & Sievert, N. W. (1973). Measuring manual dexterity with the GATB: Wooden versus plastic pegboards. Journal of Employment Counseling, 10(2), 71-77.

Kujoth, R. (1973). The validity of the GATB for the educationally deficient. Journal of Employment Counseling, 10(1), 44-48.

Lansman, M., Donaldson, G., Hunt, E., & Yantis, S. (1982). Ability factors and cognitive processes. Intelligence, 6, 347-386.

Lee, J. A., Moreno, K. E., & Sympson, J. B. (1986). The effects of mode of test administration on test performance. Educational and Psychological Measurement, 46(2), 467-474.

Manual for the General Aptitude Test Battery, Section I: Administration and scoring (Canadian ed.). (1986). Scarborough, Ontario: Nelson Canada.

Manual for the General Aptitude Test Battery, Section II: Norms, occupational aptitude pattern structure (Canadian ed.). (1983a). Scarborough, Ontario: Nelson Canada.

Manual for the General Aptitude Test Battery, Section III: Development (Canadian ed.). (1980). Scarborough, Ontario: Nelson Canada.

Manual for the General Aptitude Test Battery, Section IV: Norms, specific occupations (Canadian ed.). (1983b). Scarborough, Ontario: Nelson Canada.

McKee, L. M., & Levinson, E. M. (1990). A review of the computerized version of the Self-Directed Search. The Career Development Quarterly, 38(4), 325-333.

Moore, R. T., & Davies, J. A. (1984). Predicting GED scores on the bases of expectancy, valence, intelligence, and pretest skill levels with the disadvantaged. Educational and Psychological Measurement, 44(2), 483-490.

Reardon, R., & Loughead, T. (1988). A comparison of paper-and-pencil and computer versions of the Self-Directed Search. Journal of Counseling and Development, 67, 249-252.

Stoelting, C. (1990). A study of the construct validity of the MESA. Vocational Evaluation and Work Adjustment Bulletin, 23(3), 85-91.

Tabachnick, B. G., & Fidell, L. S. (1989). Using multivariate statistics. New York, NY: Harper Collins Publishers.

Trimmer, H., Jr., & Klein, F. (1974). Wooden versus plastic apparatus with the GATB: A restudy. Journal of Employment Counseling, 11(3), 123-126.

U. S. Employment Service, Division of Testing Staff. (1978). Ten years of USES research on the disadvantaged. The Vocational Guidance Quarterly, 26(4), 334-341.

Vevea, J. L., Clements, N. C., & Hedges, L. V. (1993). Assessing the effects of selection bias on validity data for the General Aptitude Test Battery. Journal of Applied Psychology, 78(6), 981-987.

Weber, L., King, M. F., & Pitts, G. (1973). Predicting success in schools of practical nursing. Educational and Psychological Measurement, 33(4), 943-945.

Appendix A

Informed Consent Form

LAKEHEAD UNIVERSITY
955 Oliver Road, Thunder Bay, Ontario, Canada P7B 5E1
Department of Psycholog
Telephone (807) 343-844

Informed Consent Form

1. Topic of research: The Comparison of Different Aptitude Tests and their Relationship to University Achievement

2. I, _____, consent to participate in this study on aptitude tests which investigates the differences between different aptitude tests and the relationship of aptitude scores and university achievement.

3. The researcher, Kevin Yeasting, has told me what I am supposed to do in this study. I will complete a conventional paper-and-pencil aptitude test which will last approximately one and a quarter hours. I will also complete another aptitude test which will be administered by computer and will last approximately one hour. Completion of these tests will not necessarily be in this order. In addition, there will be a period of about 3 to 4 weeks between aptitude test administrations.

4. All of my responses will be kept anonymous and confidential by the researcher. This means that no one else, except for Kevin Yeasting, will know my responses to the tests I complete.

5. I also agree to provide my final semester grades to the researcher for research purposes related to this study. Access to grades will be allowed either directly from the registrar or from myself voluntarily providing an official copy of my transcript for the semester.

6. I understand that there is no anticipated risk to myself for participation.

7. If for some reason I wish to discontinue my participation in the study once the session has begun, I am free to do so without explanation or penalty even after I have signed this consent form.

I have read the above about my participation in the study and I agree to participate in the study.

_____ _____
Signature of Participant Date

_____ _____
Signature of Witness Date

Appendix B

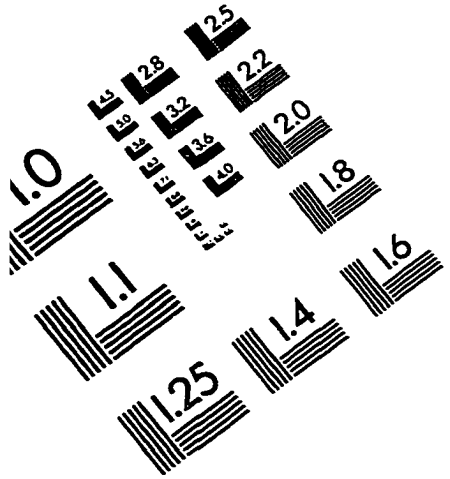
Debriefing Script

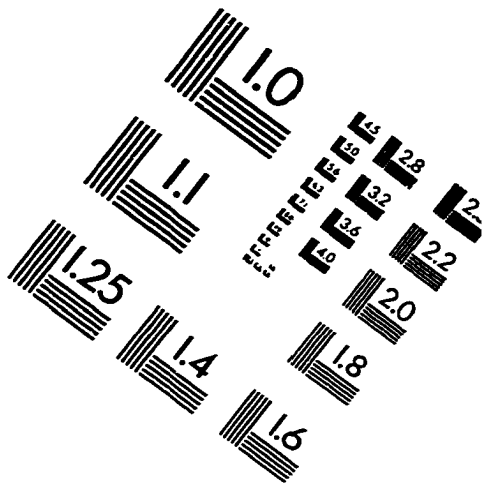
Thank-you for participating in this study. Before you leave, I want to give you some more details about the study. Generally, there were two reasons for conducting this research. First, to compare a new computer administered version of the General Aptitude Test Battery (GATB-CA) with the original paper-and-pencil version of the GATB to test for equivalence in test scores and subject response speeds. Second, to investigate whether scores from the GATB and GATB-CA can be used to predict success in subjects obtained from specialized programs of study. This research is important because it is the first to be conducted using the recently developed GATB-CA.

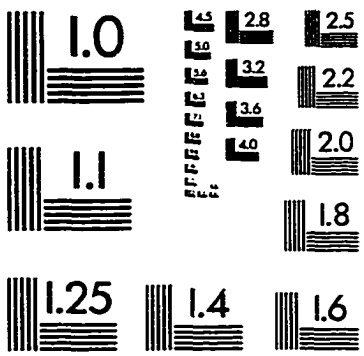
If you wish to know more about the study or obtain information regarding the overall results of the study, please leave your name, address, and telephone number with me. I will then provide the requested information upon completion of the research.

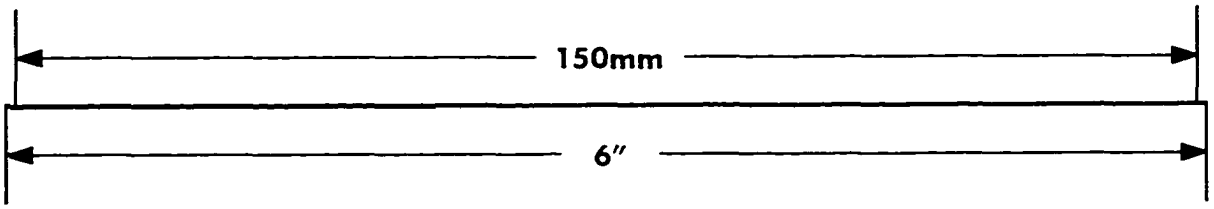
Again, thank-you very much for your participation. It was a pleasure meeting you.

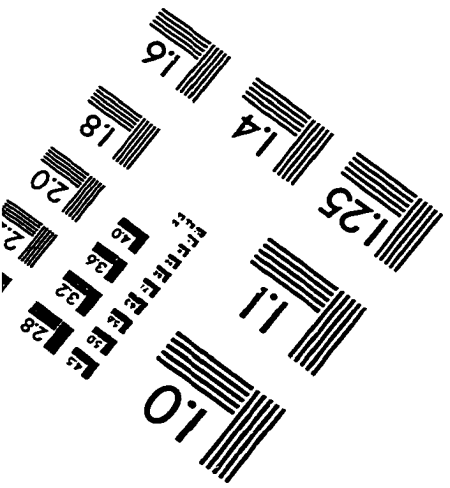
IMAGE EVALUATION TEST TARGET (QA-3)

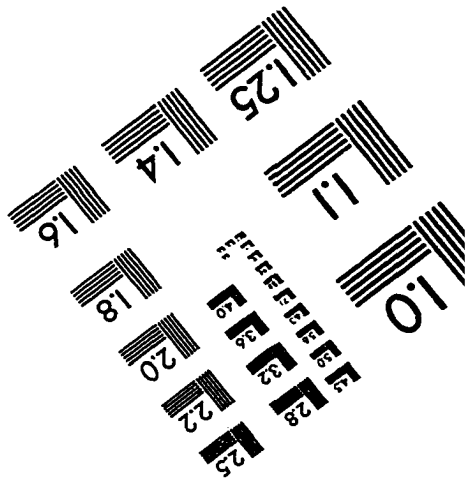












APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989