Assessing the Validity of the Taylor Visagraph II

-Lee Newton, 4th year optometry student, Sandra MacPhee, 2nd year optometry student,
Mark Swan, O.D., M.Ed., faculty advisor

Introduction:

Tests of reading ability have long been a standard element of schools and teaching institutions. Such tests are useful for documenting the normal growth and maturation of the child as a reader as well as assessing reading problems or deficiencies.

Reading is traditionally defined as the act of decoding and comprehending text. All tests of reading ability must demonstrate measurement of both components to actually evaluate one's reading. Comprehension is customarily measured by the administration of a series of questions about the test material or passage. Decoding is assessed a number of different ways; typically, it involves timing the reading of a passage, counting mistakes made, or evaluating phonation and receptive vocabulary. It is possible that evaluating oral reading may be somewhat unnatural; that is, decoding and comprehension scores after oral reading are sometimes lower than after silent reading.

An alternative model attempts to assess the decoding process by objectively monitoring eye movements made during reading. Through the use of normative data, this model converts raw eye movement data into a grade level score. The grade level score, when considered in conjunction with the comprehension score, can be used to predict one's reading level or ability.

As early as 1879, researchers have known that one's eyes do not continuously scan across the text while reading (Paulson and Goodman, 1999). Rather, the eyes make a series of short jumps, called saccades, while reading. Saccades are found to alternate with fixations, short intervals of time in which the eyes are paused on an area of text. Therefore, the act of reading can be thought of as a "stop" while the visual system processes the current material, followed by a "hop" to the succeeding area of text.

Furthermore, the work of Edmund Burke Huey (1908) showed that the number of fixations made by a reader varied as a function of text difficulty, but in no case did a one-to-one relationship exist between the number of fixations and number of words read (Paulson and Goodman, 1999). Therefore, one may be able to "process" more than one word per fixation with relatively easy text, but require *one or more* fixations per word when reading more difficult text. This information is not obtained through analysis of each particular eye movement, but rather by an average computed using the total number of eye movements and the total number of words read. It has been shown that superior readers are able to read material while requiring relatively fewer fixations than poor readers. This implies that superior readers take in more information per fixation than do poor readers.

Another parameter of reading eye movements worthy of attention is frequency of regressions. A regression is essentially the opposite of a saccade – it is a "hop" back in the right-to-left direction to reread previously read material. Regressions may occur for a

number of reasons: the reader may feel compelled to reread a section of moderate difficulty for understanding; the reader may never have developed efficient eye movements; or, the reader may never have developed adequate reading skills. Superior readers tend to make fewer regressions than poor readers.

The Taylor Visagraph II is a computerized device designed to assess one's reading level by monitoring eye movements made during reading. It consists of goggles with an internal infrared light source; these are connected by a cable to a central control box, which in turn is connected to a computer. To use the Visagraph II, a subject wears the goggles, holds a testing booklet at a comfortable reading distance, and silently reads a paragraph of text. Then the subject answers a series of 10 comprehension questions about the reading material. Using information from the subject's eye movements, the Visagraph II calculates a "Relative Efficiency" and a corresponding "Grade Equivalent" for each reading trial. The Visagraph II is expected to provide a simple, adjunct means by which to diagnose reading problems and monitor the efficacy of therapy.

The purpose of this study is to compare reading levels reported by the Visagraph II to those of other standardized reading tests, in an attempt to establish validity in measuring reading ability in such a manner. A previous study (Colby *et al.,* JAOA 1/98) assessed both the reliability and validity of the Visagraph II: the validity was measured by comparing subjects' scores on the Visagraph with scores on the reading comprehension portion of the Optometry Admissions Test (OAT). That study found that the OAT reading comprehension score correlated significantly with subjects' duration of fixations and comprehension score on the Visagraph II. However, the OAT reading comprehension score may be an artificial measure of reading in that 1) there is no assessment of the decoding process and 2) candidates are not required to read selections in their entirety.

This study utilizes other standardized reading tests, with pre-determined reliability, validity, and normative data, to determine if the results from the Visagraph II yield similar information regarding a subject's reading ability. The intent is to ascertain the validity of several standardized reading tests and to establish a procedure for comparing their results.

Methods:

Subjects:

Ten college-age subjects (5 female, average age 24y 8mo; 5 male, average age 22y 8mo) were used in this study. Subjects included were those who:

1) Had never been diagnosed with dyslexia or other reading disability.
2) Reported they do not experience any discomfort while reading or doing other types of near work.
3) Demonstrated best-corrected visual acuities of 20/20 or better, each eye, distance and near, and wore the correction for the testing procedures.
4) Demonstrated 40" or better of stereopsis.
5) Demonstrated a gradient AC/A ratio between 3/1 and 7/1.
6) Met Sheard's criterion at near.

Subjects were volunteers who received no academic or financial rewards for their involvement. In addition, each subject read and signed an "Agreement to Participate". This document outlined the purpose of the study and listed the principal faculty advisor as well as that individual's telephone number. The procedures for this study were approved by the Ferris State University Committee for the Protection of Human Subjects.

Instrumentation:

As described, the Taylor Visagraph II computer software, goggles, reading selections, and testing protocol were used in this study. The Gray Oral Reading Test (GORT), Test of Reading Comprehension (TORC), and Woodcock Diagnostic Reading Battery (WDRB) were also used.

The GORT involves subjects reading several passages orally (with examiner timing the process and recording miscues) and answering comprehension questions. The GORT reports a passage score (involving timing and accuracy of reading a selection) and a comprehension score (related to number of comprehension questions correct). It also gives an Oral Reading Quotient, which is related to the sum of the passage and comprehension scores. Grade levels are given for each measure.

The TORC is comprised of several subtests; in each, subjects must read passages and answer several questions from a test booklet. Only the subtests that contribute to the Reading Comprehension Quotient were administered and scored. This test is untimed.

The WDRB consists of 10 subtests. In this study, 4 were administered: Letter-Word Identification (receptive vocabulary), Word Attack (phonation), Reading Vocabulary (expressive vocabulary), and Passage Comprehension (deriving meaning from context, passage completion). In addition, the WDRB uses scores from these 4 subtests to compute 4 composite reading scores. Each subtest score and composite score is associated with a Grade Equivalent score and standard score. The WDRB also provides the option of further assessing "predicted standard scores" which attempt to determine overall reading level based on the results of various combinations of test scores. Since 4 subtests were administered, this option was not utilized.

Procedures:

After obtaining consent and performing the pretesting activities, subjects were introduced to the Taylor Visagraph II system. Each subject was seated comfortably in an adjustable stool and allowed to position the goggles to his or her comfort. The goggles were worn over the individual's habitual spectacle or contact lens correction. The examiner then aligned the goggles to the correct interpupillary distance. Each subject randomly selected three numbers between 1 and 10, inclusive. These numbers were used to determine the paragraphs that would be read from the College/Adult portion of the Visagraph II examination manual, which consists of ten selections. To rule out prior awareness of paragraph material, each subject was instructed to "Tell me everything you know about ____(paragraph topic)". In the event of prior awareness, which was defined as three or more correct facts, an alternate selection was chosen.

The subjects were then acclimated to the Visagraph II by performing an initial "practice" reading assessment with a less advanced selection, "Covered Wagons". Subjects were instructed that they were to read the passage silently while the Visagraph II monitored their eye movements. When they had finished, they were to answer 10 comprehension questions on the material they had just read. Subjects were given the opportunity to ask questions about the instructions and procedure. Subjects were then told "Look at the circle above the selection. Start reading when you hear a beep. Close your eyes when you are finished." Subjects were given the option of whether to answer the comprehension questions after the practice passsage, and were told that the actual tests would be similar but slightly more difficult. In a similar fashion, the three randomly selected passages from the College/Adult section of the Visagraph II were administered, along with the subsequent comprehension questions. Thus, Visagraph II trials were run four times: one initial "practice" run to acclimate the subject to the test, and three trials of grade-appropriate material. The results were then saved on the computer's internal drive and a hard copy obtained for each.

Following the Visagraph II testing, 3 other standardized reading tests (GORT, TORC, WDRB) were administered. Subjects were given the option of whether to complete all the testing at the first session or to return and finish the testing. In each case, all the appropriate Visagraph II testing was completed at the first session.

Scoring:

The Visagraph II requires minimal scoring effort from the examiner. The computer program automatically calculates all parameters from each reading trial, and also stores the answers to the comprehension questions. One must obtain a 70% score on the comprehension questions for the parameters of reading eye movements to be valid. For the sake of consistency in computing the scores from the Visagraph II testing, only the first two grade-appropriate trials that satisfied the 70% or better comprehension requirement were used. Then, the arithmetic mean was computed for the scores in these two trials scores to give "average" Visagraph II scores.

The Visagraph II calculates many parameters of reading eye movements; however, the following are used to determine grade level:

1) Fixation per 100 words – Number of times the eyes stop moving per 100 words read.
2) Regression per 100 words – Number of times eyes move back to the left per 100 words read.
3) Reading Rate with Comprehension – Reading rate for selection, excluding first and last lines of selection.

"Relative Efficiency" is a calculation using these three measures. From this, a "Grade Level" is provided by the Visagraph II.

"Grade Level" can also be referenced from a table of normative data. Possible discrepancies can develop from evaluating data in this manner: The Visagraph II gives an exact Grade Level based on Relative Efficiency, and the normative table consists of Grade Level based on discrete intervals of Relative Efficiency. It is not possible to interpolate data due to the non-linear relationship. For example, if a Relative Efficiency of 10.77 corresponds to a Grade Level of 17.0 and a Relative Efficiency of 13.48 corresponds to a Grade Level of 18.0, a Relative Efficiency of 12.0 would logically imply a Grade Level of near 17.5. However, since the relationship is non-linear and there are no levels of Relative Efficiency between 10.77 and 13.48 in the table, one would still have to use 17.0 (the last level on the table actually achieved). Therefore, since the Visagraph II provides an *exact* Grade Level for each testing session (it essentially "interpolates" the non-linear relationship automatically), it is more accurate to use an average of Grade Levels rather than to average the Relative Efficiencies and *then* reference a Grade Level. The former method, averaging of Grade Levels, is used in this study.

The GORT, TORC, and WDRB were all scored in accordance with their respective scoring manuals. In computing the data, grade levels were often "averaged" arithmetically in accordance with the base-10 numeric system. Although this seems to work well numerically, a potential inconsistency is that there are ten digits in the Arabic system but traditionally 9 months in a standard school year. Typically, the zero is avoided since it is impossible to be in the "zeroth" month of a school year. Ultimately, this phenomenon should produce little, if any, discrepancy in the results.

Even though one may expect reasonable correlations to be obtained through multiple measures of reporting scores, various idiosyncrasies of the tests may have precluded doing so in a reliable fashion. For example, the best score on the "Passage" portion of the GORT is a Grade Equivalent of ">12.9". Nine of the 10 subjects attained a "Passage" Grade Equivalent of ">12.9". It is not possible to analyze numerical data presented in such a manner. If this score is ascribed a singular real number for purposes of data analysis, it is not possible to determine *to what extent* each score exceeded 12.9. In this study, scores of ">12.9" were reported as "13.0" to facilitate statistical manipulation of data.

Results:

Each subject's score on every test and subtest is displayed in Appendix A. Please note that Grade Equivalents are ignored in all but the final, average Visagraph II score. Although each measured eye movement parameter has a corresponding Grade Equivalent score, it is the *actual measured parameter* that is used to compute the Relative Efficiency score (and, in turn, its corresponding Grade Equivalent score). Therefore, other measures of Grade Equivalent are superfluous. Also, the Visagraph II reports eye movement results separately for the left and right eyes. In no case was a significant discrepancy obtained; however, in the event of a difference, the arithmetic mean was used.

Correlation coefficients were computed between each measure and all other measures. Correlation coefficients range from –1 to +1, with -1 indicating perfect

negative correlation, 0 representing absolute lack of correlation, and +1 representing perfect positive correlation. Correlation does not imply causation. Please note that some apparently strong correlations are meaningless as they are merely different computations of the same data. For example, the correlation between rows 6 & 8 is artificially strong – the measures come from the same data. Also, for any correlation to be significant at the P<5% level (using 10-2=8 degrees of freedom), the coefficient must be at least 0.632. Otherwise, there is insufficient data to differentiate this from a chance occurrence. Correlation coefficients are displayed in Table 1.

Descriptive statistics, such as mean, median, and standard deviation, were computed for each test using the scores from all subjects. These data are displayed in Table 2.

Scores for each test were also arranged in decreasing order of proficiency by subject. These data are displayed in Table 3.

Discussion:

The fact that Relative Efficiency correlated well between trials 1 & 2 (0.900) shows that the Visagraph II does indeed provide a repeatable assessment of reading parameters. Relative Efficiency is a composite of the subjects' rates of reading, fixations, and regressions. Average Words-per-Minute correlated well with Average Fixations/100 words (-0.918); this implies that these two measures provide a valid assessment of reading eye movements. Average Words-per-Minute was found to correlate less well, however, with Average Regressions/100 words (-0.592). This did not meet our threshold for significance at the P<5% level (0.632).

Interestingly, comprehension scores between trials 1 & 2 exhibited a significant negative correlation (-0.703). The reason for this is not known. Perhaps it pertains to the fact that some Visagraph II trials were discarded and only those with a comprehension score greater than or equal to 70% were used; a conceivably "biased" representation of the data.

The WDRB measures all had high intra-test correlation; most exceeded the criterion for significance.

The GORT Comprehension and Passage scores did not correlate well (-0.124). However, the Comprehension score correlated better with the Oral Reading Quotient (ORQ) (0.808) than the Passage score did with the ORQ (0.324).

Because essentially all subjects attained Grade Level scores on the TORC subtests of ">12.0", only the composite score (Reading Comprehension Quotient {RCQ}) was used. This measure was felt to more accurately differentiate the achievement of the subject population.

The Visagraph II scores correlated poorly with the measures given by the other tests. The only correlations near statistical significance were between the WDRB

Reading Vocabulary subtest (WDRB RV) and Visagraph II average words-per-minute (0.628) and WDRB RV and Visagraph II average Relative Efficiency (0.593). The Visagraph Average Relative Efficiency had a 0.212 to 0.400 correlation between composite reading measures on the WDRB; a –0.036 to 0.168 correlation with GORT measures; and a 0.216 correlation with the TORC RCQ[1]

The WDRB composite measures correlated poorly with the GORT; the highest correlation was 0.615 between the WDRB Total Reading and GORT ORQ. The WDRB correlated more highly with the TORC; significant correlations were obtained between the TORC and the WDRB Letter-Word Identification (0.743), Reading Vocabulary (0.659), Total Reading (0.669), and Basic Reading Skills (0.722).

The TORC correlated marginally with the GORT. The highest correlation was between the GORT ORQ and the TORC RCQ (0.481); shy of statistical significance.

Table 2 provides descriptive statistics for each row in Appendix A. From this representation, one can draw conclusions about the average Grade Level scores as well as the range of scores. The average Grade Equivalent given by the Visagraph II was 9.33, with a median of 8.85 and a range of 5.5 to 15.7. For the GORT Average Grade Equivalent, the mean was 11.83, with a median of 12.6 and a range of 7.7 to 13. The WDRB composites provided mean Grade Equivalents of between 15.21 and 16.9, with more closely spaced ranges (minimum Grade Equivalent was 11.3 in any case).

An interesting relationship was demonstrated between the GORT ORQ and TORC RCQ – these measures paralleled almost exactly. The GORT ORQ had a mean of 107.2, median of 109, and a standard deviation of 14.30. The TORC RCQ had a mean of 106.7, median of 112, and a standard deviation of 16.31. Although the TORC does not provide a Grade Equivalent, sufficient evidence existed to show that the two measures provided similar results (despite a non-significant correlation of 0.481).

Table 3, a relative ranking of subjects based on proficiency for each row of data, provides a qualitative view of score distributions. Within the Visagraph II scores, it is evident that the same subjects performed the best in all categories. The weakest relationship was with Regressions. Since all subjects scored 70% or better on the Visagraph comprehension questions, these data were clustered sufficiently enough as to preclude inferring any trends.

The Grade Equivalent scores from the WDRB and GORT did not adequately differentiate subjects; most scored the maximum of 13 or 16.9. Between the GORT ORQ, TORC RCQ, and Visagraph II Average Relative Efficiency, the scores for the most part had continuity in their distribution. For example, subjects were in roughly the same one-third of the distribution in each case. There was one glaring exception – one subject

[1] Since this preliminary study has a small sample size, the raw data correlations between each subtest are occasionally listed as a range, instead of calculating average scores of the small sample and then testing for correlation.

scored second-best on the Visagraph II, worst on the GORT ORQ, and third-worst on the TORC RCQ.

Conclusion:

Possible sources of error or confusion in this study include utilizing standardized tests whose scoring was not normed for the particular population under study. The WDRB, GORT, and TORC all were normed for individuals through 18 years old or through the twelfth grade, while the Visagraph II is appropriate through adulthood and provides Grade Equivalent through the 18th grade. Nevertheless, we feel that the test sample used in this study (>18 years of age) rendered valid data for the following reasons: 1) Grade level is a more reliable method than age for reporting reading ability (for instance, there is no standard reading level of a 40-year old), and 2) the population of subjects in this study met our entrance criteria and were educated individuals who demonstrated (on most measures) a reading level at or near appropriate grade levels.

The absence of a significant correlation between the Visagraph II achievement measures and those of other reading tests is insufficient to disprove validity – mainly due to the complexity of the reading process. It would be unwise to assume that the mechanical monitoring of eye movements provides a *complete* analysis of the interaction between the eyes and brain during reading. However, this method has shown that it arrives at realistic approximations of reading aptitude based on the mechanical data. The Visagraph II provides a similar range of scores as most of the other measures, generally involves subjects' scores in the same positions of relative achievement as the other tests, and provides information not only about comprehension but also about eye movements involved with the reading act. Furthermore, it is an easy and useful test to perform to either establish the presence or absence of a reading deficiency or to monitor the efficacy of any intervention. Future studies are planned to assess adults and children with reading proficiency levels that are more appropriate for the difficulty levels of the tests used.

References

Colby, Dirk, O.D.; Laukkaken, Hannu R., O.D., M.Ed.; Yolton, Robert L., Ph.D., O.D. "Use of the Taylor Visagraph II system to evaluate eye movements made during reading." *Journal of the American Optometric Association* 1994: 69 (1) pp. 22-32.

Dwyer, Edward J.; West, Russell F. "Demystifying 'Speed Reading': A Practical Approach for Increasing Rate." *Forum for Reading* Spring/Summer 1990: 21(2) pp. 68-74.

Grisham, David; Simons, Herbert. "Perspectives on Reading Disabilities", Chapter 22 in *Diagnosis and Management of Learning-Related Vision Disorders*, pp. 518-542.

Paulson, Eric J.; Goodman, Kenneth S. "Influential Studies in Eye-Movement Research." Reading Online, http://www.readingonline.org/research/eyemove.html, 1994.

"Statistics," Martin Sternstein, Ph.D., Barron's Educational Series, Inc., 1994

Test of Reading Comprehension, Examiner's Manual.

Visagraph II Administration and Scoring Manual, Taylor and Associates.

Woodcock Diagnostic Reading Battery, Examiner's Manual.

Wiederholt, J. Lee; Bryant, Brian R. Gray Oral Reading Test Examiner's Manual

Subject		1	2	3	4	5
Row 1	VISA 1 R.E.	1.91	2.44	4.92	1.53	1.06
Row 2	VISA 1 COMP.	8	7	9	9	8
Row 3	VISA 2 R.E.	1.71	2.38	9.98	1.51	1.45
Row 4	VISA 2 COMP.	9	10	8	7	9
Row 5	VISA AVE. WPM	234	250.5	474	210.5	180.5
Row 6	VISA AVE. F/100	112.5	99.5	62.25	113.5	124
Row 7	VISA AVE. R/100	16.75	4.5	5	23.25	21.75
Row 8	VISA AVE. R.E.	1.81	2.41	7.45	1.52	1.255
Row 9	GE	9.6	13	15.7	7.65	5.9
Row 10	WDRB L-W I. GE	16.9	16.9	16.9	16.8	16.9
Row 11	WDRB W.A. GE	16.9	11.9	16.9	11.9	16.9
Row 12	WDRB R.V. GE	15.9	13.9	16.9	13.2	14.8
Row 13	WDRB P.C. GE	15.6	16.9	16.9	14.2	16.9
Row 14	WDRB T.R. (1-4) GE	16.3	14.9	16.9	14	16.4
Row 15	WDRB B.R. (1,4) GE	16.3	16.9	16.9	15.5	16.9
Row 16	WDRB B.R.S. (1,2) GE	16.9	14.4	16.9	14.4	16.9
Row 17	WDRB R.C. (3,4) GE	15.8	15.4	16.9	13.7	15.9
Row 18	GORT COMP. GE	13	2.4	10.3	11.9	13
Row 19	GORT PASS. GE	13	13	13	12.4	13
Row 20	GORT ORQ	115	82	109	94	121
Row 21	TORC RCQ	125	100	112	105	112

VISA = Taylor Visagraph
WPM = Words per Minute
GE = Grade Equivalent
F/100 = Fixations per 100 words
R/100 = Regressions per 100 words
COMP = Comprehension

WDRB = Woodcock Diagnostic Reading Batt
L-W I = Letter-Word Identification
W.A. = Word Attack
R.V. = Reading Vocabulary
P.C. = Passage Comprehension
T.R. = Total Reading
B.R. = Broad Reading
B.R.S. = Basic Reading Skills
R.C. = Reading Comprehension

Appendix A: Representation of all scores of all subjects

6	7	8	9	10
1.82	1.3	2.98	2.06	1.23
9	9	8	8	8
1.73	1.37	1.98	1.22	1.2
7	7	7	8	7
210.5	166	264	196.5	184.5
107.25	108	92.25	114.25	128.75
11	16.5	7	9	22.75
1.775	1.335	2.48	1.64	1.215
9.4	6.35	11.9	8.3	5.5
16.9	10.7	16.8	16.8	16.9
14.4	11.9	16.9	14.4	16.9
15.9	12.5	14.8	14.8	14.8
16.9	12	13	16.9	16.9
16	11.8	15.4	15.7	16.4
16.9	11.4	14.9	16.9	16.9
15.7	11.3	16.9	15.6	16.9
16.4	12.3	13.9	15.9	15.9
13	8.2	9.1	13	13
13	13	13	13	13
130	94	103	115	109
113	72	122	118	88

ery

GORT = Gray Oral Reading Test
Comp. = Comprehension
Pass. = Passage (regarding timing and accuracy)
ORQ = Oral Reading (Quotient

TORC = Test of Reading Comprehension
RCQ = Reading Comprehension Quotient

	Row 1	*Row 2*	*Row 3*	*Row 4*	*Row 5*	*Row 6*	*Row 7*	*Row 8*
Row 1	1							
Row 2	0.108751	1						
Row 3	0.900223	0.299192	1					
Row 4	0.0985	-0.70306	0.097706	1				
Row 5	0.965254	0.183702	0.96737	0.132996	1			
Row 6	-0.95511	-0.238799	-0.875582	-0.04325	-0.91851	1		
Row 7	-0.75309	0.223191	-0.494489	-0.263503	-0.592446	0.75437	1	
Row 8	0.950254	0.247331	0.991066	0.100062	0.987644	-0.918878	-0.584392	1
Row 9	0.924437	-0.093375	0.756247	0.303912	0.869739	-0.911295	-0.848072	0.824119
Row 10	0.25396	-0.367112	0.148948	0.297616	0.285042	-0.033865	-0.130128	0.184334
Row 11	0.272954	-0.107486	0.278779	0.021974	0.320723	-0.088424	0.02459	0.283024
Row 12	0.5832	0.043954	0.579861	0.157894	0.628211	-0.425256	-0.359698	0.593425
Row 13	0.134129	-0.32921	0.236093	0.478993	0.212281	0.056797	-0.133802	0.209968
Row 14	0.351627	-0.25415	0.352033	0.2851	0.413735	-0.118196	-0.149828	0.359522
Row 15	0.21059	-0.381102	0.206268	0.424182	0.268884	0.014366	-0.144017	0.212053
Row 16	0.308156	-0.25952	0.253248	0.165821	0.354165	-0.074858	-0.054195	0.275543
Row 17	0.343712	-0.198378	0.41163	0.387506	0.416065	-0.145311	-0.242393	0.399729
Row 18	-0.272432	0.428866	-0.139279	-0.403818	-0.188165	0.394501	0.546482	-0.183076
Row 19	0.182925	-0.364405	0.124137	0.287348	0.105137	-0.136448	-0.452554	0.144828
Row 20	-0.089761	0.26944	-0.011286	-0.161013	-0.048757	0.186808	0.141456	-0.035567
Row 21	0.357202	-0.18269	0.148952	0.270519	0.313457	-0.214854	-0.315188	0.21596

Table 1: Correlation coefficients between measures of Appendix A.

Row 9	*Row 10*	*Row 11*	*Row 12*	*Row 13*	*Row 14*	*Row 15*	*Row 16*	*Row 17*
1								
0.316081	1							
0.106684	0.464896	1						
0.508509	0.612479	0.747142	1					
0.155432	0.688315	0.318972	0.609507	1				
0.297355	0.838056	0.792951	0.881349	0.779549	1			
0.256176	0.920299	0.428513	0.666445	0.917119	0.881221	1		
0.237528	0.8318	0.878022	0.798723	0.56553	0.947801	0.762448	1	
0.321165	0.724308	0.55419	0.852869	0.933516	0.912088	0.901839	0.73424	1
-0.486127	0.253332	0.512179	0.383653	0.209157	0.419336	0.255601	0.460598	0.319739
0.177153	-0.09907	0.458831	0.415466	0.265365	0.319908	0.090379	0.229671	0.369569
-0.207255	0.329891	0.614	0.661649	0.440463	0.614991	0.421139	0.565292	0.595377
0.412869	0.74267	0.508565	0.658722	0.330301	0.668983	0.590541	0.722477	0.512641

Row 18	*Row 19*	*Row 20*	*Row 21*
1			
-0.12403	1		
0.808321	0.324408	1	
0.328859	0.036623	0.480603	1

Row1		*Row2*		*Row3*		*Row4*	
Mean	2.125	Mean	8.3	Mean	2.453	Mean	7.9
Standard E	0.361412	Standard E	0.213437	Standard E	0.844051	Standard E	0.34801
Median	1.865	Median	8	Median	1.61	Median	7.5
Mode	#N/A	Mode	8	Mode	#N/A	Mode	7
Standard D	1.142884	Standard D	0.674949	Standard D	2.669124	Standard D	1.100505
Sample Va	1.306183	Sample Va	0.455556	Sample Va	7.124223	Sample Va	1.211111
Kurtosis	3.860611	Kurtosis	-0.282995	Kurtosis	9.500662	Kurtosis	-0.521601
Skewness	1.837996	Skewness	-0.433637	Skewness	3.057412	Skewness	0.862823
Range	3.86	Range	2	Range	8.78	Range	3
Minimum	1.06	Minimum	7	Minimum	1.2	Minimum	7
Maximum	4.92	Maximum	9	Maximum	9.98	Maximum	10
Sum	21.25	Sum	83	Sum	24.53	Sum	79
Count	10	Count	10	Count	10	Count	10

Row5		*Row6*		*Row7*		*Row8*	
Mean	237.1	Mean	106.225	Mean	13.75	Mean	2.289
Standard E	28.11156	Standard E	5.924109	Standard E	2.33244	Standard E	0.589972
Median	210.5	Median	110.25	Median	13.75	Median	1.7075
Mode	210.5	Mode	#N/A	Mode	#N/A	Mode	#N/A
Standard D	88.89657	Standard D	18.73368	Standard D	7.375824	Standard D	1.865654
Sample Va	7902.6	Sample Va	350.9507	Sample Va	54.40278	Sample Va	3.480666
Kurtosis	6.834921	Kurtosis	3.006712	Kurtosis	-1.7756	Kurtosis	8.513169
Skewness	2.476464	Skewness	-1.476053	Skewness	0.086283	Skewness	2.850839
Range	308	Range	66.5	Range	18.75	Range	6.235
Minimum	166	Minimum	62.25	Minimum	4.5	Minimum	1.215
Maximum	474	Maximum	128.75	Maximum	23.25	Maximum	7.45
Sum	2371	Sum	1062.25	Sum	137.5	Sum	22.89
Count	10	Count	10	Count	10	Count	10

Row9		*Row10*		*Row11*		*Row12*	
Mean	9.33	Mean	16.25	Mean	14.9	Mean	14.75
Standard E	1.053702	Standard E	0.616847	Standard E	0.726483	Standard E	0.414528
Median	8.85	Median	16.9	Median	15.65	Median	14.8
Mode	#N/A	Mode	16.9	Mode	16.9	Mode	14.8
Standard D	3.3321	Standard D	1.950641	Standard D	2.297341	Standard D	1.310852
Sample Va	11.10289	Sample Va	3.805	Sample Va	5.277778	Sample Va	1.718333
Kurtosis	-0.244052	Kurtosis	9.98392	Kurtosis	-1.807479	Kurtosis	-0.141624
Skewness	0.736068	Skewness	-3.158828	Skewness	-0.472514	Skewness	-0.169443
Range	10.2	Range	6.2	Range	5	Range	4.4
Minimum	5.5	Minimum	10.7	Minimum	11.9	Minimum	12.5
Maximum	15.7	Maximum	16.9	Maximum	16.9	Maximum	16.9
Sum	93.3	Sum	162.5	Sum	149	Sum	147.5
Count	10	Count	10	Count	10	Count	10

Table 2: Descriptive statistics of all subjects' scores on each individual measure

Row13		*Row14*		*Row15*		*Row16*	
Mean	15.62	Mean	15.38	Mean	15.95	Mean	15.59
Standard E	0.594568	Standard E	0.479305	Standard E	0.553223	Standard E	0.575702
Median	16.9	Median	15.85	Median	16.9	Median	16.3
Mode	16.9	Mode	16.4	Mode	16.9	Mode	16.9
Standard D	1.880189	Standard D	1.515696	Standard D	1.749444	Standard D	1.820531
Sample Va	3.535111	Sample Va	2.297333	Sample Va	3.060556	Sample Va	3.314333
Kurtosis	-0.16497	Kurtosis	2.952086	Kurtosis	5.79057	Kurtosis	2.762721
Skewness	-1.16397	Skewness	-1.672756	Skewness	-2.341277	Skewness	-1.636189
Range	4.9	Range	5.1	Range	5.5	Range	5.6
Minimum	12	Minimum	11.8	Minimum	11.4	Minimum	11.3
Maximum	16.9	Maximum	16.9	Maximum	16.9	Maximum	16.9
Sum	156.2	Sum	153.8	Sum	159.5	Sum	155.9
Count	10	Count	10	Count	10	Count	10

Row17		*Row18*		*Row19*		*Row20*	
Mean	15.21	Mean	10.69	Mean	12.94	Mean	107.2
Standard E	0.453982	Standard E	1.083969	Standard E	0.06	Standard E	4.521062
Median	15.85	Median	12.45	Median	13	Median	109
Mode	15.9	Mode	13	Mode	13	Mode	115
Standard D	1.435618	Standard D	3.427811	Standard D	0.189737	Standard D	14.29685
Sample Va	2.061	Sample Va	11.74989	Sample Va	0.036	Sample Va	204.4
Kurtosis	0.327058	Kurtosis	3.426507	Kurtosis	10	Kurtosis	-0.273832
Skewness	-1.061964	Skewness	-1.813291	Skewness	-3.162278	Skewness	-0.248847
Range	4.6	Range	10.6	Range	0.6	Range	48
Minimum	12.3	Minimum	2.4	Minimum	12.4	Minimum	82
Maximum	16.9	Maximum	13	Maximum	13	Maximum	130
Sum	152.1	Sum	106.9	Sum	129.4	Sum	1072
Count	10	Count	10	Count	10	Count	10

Row21	
Mean	106.7
Standard E	5.157626
Median	112
Mode	112
Standard D	16.30985
Sample Va	266.0111
Kurtosis	1.090844
Skewness	-1.18209
Range	53
Minimum	72
Maximum	125
Sum	1067
Count	10

Point	*Row1*	*Point*	*Row2*	*Point*	*Row3*	*Point*	*Row4*
3	4.92	3	9	3	9.98	2	10
8	2.98	4	9	2	2.38	1	9
2	2.44	6	9	8	1.98	5	9
9	2.06	7	9	6	1.73	3	8
1	1.91	1	8	1	1.71	9	8
6	1.82	5	8	4	1.51	4	7
4	1.53	8	8	5	1.45	6	7
7	1.3	9	8	7	1.37	7	7
10	1.23	10	8	9	1.22	8	7
5	1.06	2	7	10	1.2	10	7

Point	*Row5*	*Point*	*Row6*	*Point*	*Row7*	*Point*	*Row8*
3	474	10	128.75	4	23.25	3	7.45
8	264	5	124	10	22.75	8	2.48
2	250.5	9	114.25	5	21.75	2	2.41
1	234	4	113.5	1	16.75	1	1.81
4	210.5	1	112.5	7	16.5	6	1.775
6	210.5	7	108	6	11	9	1.64
9	196.5	6	107.25	9	9	4	1.52
10	184.5	2	99.5	8	7	7	1.335
5	180.5	8	92.25	3	5	5	1.255
7	166	3	62.25	2	4.5	10	1.215

Point	*Row9*	*Point*	*Row10*	*Point*	*Row11*	*Point*	*Row12*
3	15.7	1	16.9	1	16.9	3	16.9
2	13	2	16.9	3	16.9	1	15.9
8	11.9	3	16.9	5	16.9	6	15.9
1	9.6	5	16.9	8	16.9	5	14.8
6	9.4	6	16.9	10	16.9	8	14.8
9	8.3	10	16.9	6	14.4	9	14.8
4	7.65	4	16.8	9	14.4	10	14.8
7	6.35	8	16.8	2	11.9	2	13.9
5	5.9	9	16.8	4	11.9	4	13.2
10	5.5	7	10.7	7	11.9	7	12.5

Point	*Row13*	*Point*	*Row14*	*Point*	*Row15*	*Point*	*Row16*
2	16.9	3	16.9	2	16.9	1	16.9
3	16.9	5	16.4	3	16.9	3	16.9
5	16.9	10	16.4	5	16.9	5	16.9
6	16.9	1	16.3	6	16.9	8	16.9
9	16.9	6	16	9	16.9	10	16.9
10	16.9	9	15.7	10	16.9	6	15.7
1	15.6	8	15.4	1	16.3	9	15.6
4	14.2	2	14.9	4	15.5	2	14.4
8	13	4	14	8	14.9	4	14.4
7	12	7	11.8	7	11.4	7	11.3

Table 3: Rank order of each subject's score on each particular test item.
"Point" refers to subject number; t "Row" refers to scores on test items in that row.

Point	*Row17*	*Point*	*Row18*	*Point*	*Row19*	*Point*	*Row20*
3	16.9	1	13	1	13	6	130
6	16.4	5	13	2	13	5	121
5	15.9	6	13	3	13	1	115
9	15.9	9	13	5	13	9	115
10	15.9	10	13	6	13	3	109
1	15.8	4	11.9	7	13	10	109
2	15.4	3	10.3	8	13	8	103
8	13.9	8	9.1	9	13	4	94
4	13.7	7	8.2	10	13	7	94
7	12.3	2	2.4	4	12.4	2	82

Point	*Row21*
1	125
8	122
9	118
6	113
3	112
5	112
4	105
2	100
10	88
7	72