Title: Correlating the Gates-MacGinitie Reading Tests with the University Entrance English Examinations Held in Taiwan

Author: Tzung-yu Cheng

Date Completed: May 21, 2013

INTRODUCTION

The universities in Taiwan recruit students through two national Joint College Entrance examinations (JCEE): General Scholastic Ability Test (GSAT) and Advanced Subjects Test (AST) (CEEC, 2013). The GSAT contains five subjects: Chinese, English, Mathematics, Social Studies, and Science. Each subject area is coded with a scaled score ranging from 0 to 15. Ten subjects are included in the AST: Chinese, English, Mathematics for science and engineering majors, Physics, Chemistry, Biology, Mathematics for humanities and social science majors Geography, History, and Civics. Each subject test is assigned a score ranging from 0 to 100 points. In the English examinations, a set of completely new questions is assigned to the students during every conduct of the test. The validity of the contained test questions wholly relies on the judgment and expertise of the test writers. The AST English test has been reported as a valid tool (LTTC, 2003). Nevertheless, we aimed to secure more credible evidence and further validation of the exams using different well-recognized tests. Hence, an attempt was made by correlating the English tests of the AST and the GSAT held in 2012 with a well-recognized standardized reading test for American school students, the Gates-MacGinitie Reading Tests (GMRT). The study used the GMRT as a great number of high school graduates were assigned English-language discipline-specific textbook study after entering universities (Cheng, 2010a; Cheng, 2010b). These textbooks were written for native English speakers who "ideally" should score a reading grade level of at least 12 or above (Cheng, 2010a; Singer & Donlan, 1989).

Along with this study, attempts were also made to equate both the AST and the GSAT English scores with the grade equivalents (GE) derived from the GMRT raw scores. Grade equivalents are expressed in grades and months such as 5.2, the second month of the fifth Grade (Lipson & Wixson, 1991). They are often employed in standardized norm-referenced group survey tests, such as the Gates-MacGinitie Reading Tests (GMRT) (MacGinitie, MacGinitie, Maria & Dreyer, 2002), criterion-referenced tests (Slossen, 1988), and Informal Reading Inventory (IRI) (Burns & Roe, 1989). This research is expected to pave the way to large-group studies for validating and depicting the efficacy of the language exams developed in Taiwan and in Asia.

RELATED LITERATURE

Validity is one of the most important considerations in developing and evaluating language tests. Various approaches can be used to validate a test: experts' assessment for the item-objective congruence, comparing the test scores to the

students' semester grades for predictive validity, and testing and retesting the students for internal consistency (Walt & Steyn, 2008). Yet, a very common approach to evaluate tests is through correlational research between or among different existing language tests; for instance, the Slossen Oral Reading Test was correlated with the Standard Oral Reading Paragraphs and the Peabody Individual Achievement Test of Reading Recognition (Slossen, 1988). And the Gates-MacGinitie Reading Tests were correlated with the verbal or English sections in Preliminary Scholastic Assessment Test, Scholastic Assessment Tests, American College Testing Program, and grade point average (GPAs) (Lipson & Wixson, 1991; MacGinitie, MacGinitie, Maria & Dreyer, 2002).

Test of English as a Foreign Language (TOEFL) and the International English Language Testing System (IELTS) are among well-known international English tests. An unvarying feature of these tests is that every time an entirely new set of questions is posed to the students. Judgment and expertise of the test writers should prevail to increase the content validity of the test questions. In contrast, many standardized norm-referenced reading tests or criterion-referenced tests are like an IQ test that has only one or two forms of a test. Each form consists of a fixed set of questions that is used and reused on different groups of targeted population over a long period of time or until a new edition is published to reflect new reading concepts. Each question is selected through rigorous procedures by field experts and tested statistically for reliability and content validity. Before release to the students, each test is tested and retested on a target population for possible corrections in the questions and establishing the norms which may contain Raw Scores, National Stanines, Normal Curve Equivalents (NCEs), National Percentile Ranks, Grade Equivalents (GE), and Extended Scale Scores. Hence, the money and manpower involved are usually tremendous. The information obtained from the tests, serves as an important basis for the following decision making (Ekwall & Shanker, 1988; MacGinitie, MacGinitie, Maria & Dreyer, 2000; Lipson & Wixson, 1991):

1. Evaluating the effectiveness of instructional programs;
2. Making decisions about grouping students;
3. Planning instructional emphases;
4. Locating students who are ready to work with more advanced materials;
5. Deciding the levels of instructional materials to be assigned to new students;
6. Selecting students for further individual diagnosis and special instruction;
7. Communicating to the students about their progress in reading;
8. Reporting to parents and the community.

Why Was the Gates-MacGinitie Reading Test Chosen for This Study?

In the early 20th century in the United States, psychologists and reading

specialists initiated a two-track movement using scientific methods to explore reading problems. In 1914, Thorndike developed the first norm-referenced group test of reading ability. In 1915, Williams S. Gray published an oral reading test, which led to the diagnostic movement in reading and to an emphasis on remediation (Ekwall & Shanker, 1988; Lipson & Wixson, 1991). In 1926, Arthur Gates published the Gates Silent Reading Test and the Gates Primary Reading Tests, two of the widely used reading tests. The Gates-MacGinitie Reading Tests continued the long tradition of reading tests by Arthur Gates since that time. Over the years, the Gates-MacGinitie Reading Tests have been improved and revised to reflect new concepts in reading and to establish new national norm (MacGinitie, MacGinitie, Maria & Dreyer, 2000).

The Gates-MacGinitie Reading Tests have been used at national level in the United States by school districts, classroom teachers, doctoral students, researchers, reading specialists, and in national studies sponsored by U.S. Department of Education (Carpenter & Paris, 2005; Cook, Gerber & Semmel, 1997; Drummond, Chinen, Duncan, Miller, Fryer, Zmach & Culp, 2011; Fisher, 2001; Gilbert, 2009; Johnson & McCabe, 2003, Lipson & Wixson, 1991; Nelson & Stage, 2007; Paris & Associates, 2004; Rowe, Ozuru, O'Reilly & McNamara, 2008; Tatum, 2004; Tilstra, McMaster, Broek, Kendeou & Rapp, 2009). In the EFL context, Cheng (2009, 2010) also used the Level 7/9 of the GMRT to investigate Taiwanese university students' vocabulary grade equivalents and reading grade equivalents.

The current Fourth edition contains the following grade levels: PR (Pre-Reading), BR (Beginning Reading), Level 1, Level 2, Level 3, Level 4, Level 5, Level 6, Level 7/9, Level 10/12, and AR (Adult Reading). Levels 2 through AR have two forms, Form S and Form T, for test and retest. Levels 3 through AR comprise two subtests each: Vocabulary and Comprehension.

The Vocabulary subtest measures reading vocabulary by asking students to choose one word or phrase that conveys the nearest meaning to the given word or phrase. The subtest contains 45 questions. Each word is presented in a brief context frame with five choices. This test has a time limit of 20 minutes to complete the subtest. The vocabulary test words are of general usefulness and not obscured or specialized words. Many vocabulary questions include one or more of three different types of wrong answers: visual similarity, miscue, and association (MacGinitie, MacGinitie, Maria & Dreyer, 2002).

The Comprehension subtest measures the ability of students to read and understand different types of prose. The subtest contains 48 questions. Each comprehension question is presented with four choices. The time given to complete the subtest is 35 minutes. The comprehension passages consist of a mixture of fiction, social studies, natural sciences and humanities. The passage type includes narratives,

expository and setting (Lipson & Wixson, 1991; MacGinitie, MacGinitie, Maria & Dreyer, 2002). It contains inferential and literal questions equally. The passages are selected from varied authorship and not from very familiar topics or from books or other materials that are currently very popular, or used in many classrooms, or likely to have been read by many students (Lipson & Wixson, 1991; MacGinitie, MacGinitie, Maria & Dreyer, 2002). Females and males of various ethnic groups are equally represented in test content.

In establishing the national norm for the entire tests, about 65,000 students studying in both public and private schools from all parts of the country were tested in the fall of 1998 and spring of 1999 for the Fourth Edition. Raw scores were converted into national stanines, normal curve equivalents (NCEs), national percentile ranks, grade equivalents (GE), and extended scale scores (MacGinitie, MacGinitie, Maria & Dreyer, 2002).

Johnson and McCabe (2003) reviewed the Fourth Edition of the Gates-MacGinitie Reading Tests. They pointed out that the GMRT showed strong total test and subtest internal consistency levels, ranging from 0.88 to 0.90. The significant statistic figures are listed as follows:

1. Coefficient values were at or above 0.90 for all test materials.
2. Alternate form correlations for the total tests were at or above 0.90.
3. Alternate form correlations for the subtests ranged from 0.74 to 0.92.
4. Total test coefficient values were at and above 0.88.
5. Test-re-test reliability had been reported as above 0.88.

In the review of the Fourth Edition, Johnson and McCabe (2003) affirmed strong evidence for test validity. They stated that the content validity of the GMRT is supported through an extensive test development process, and scores are reported to correlate well with the scores of similar measures such as the Standard Achievement Test. MacGinitie et al. (2002) also pointed out that the correlations between the Third and Fourth Editions were very high, ranging from 0.91 to 0.93 and the design of the two editions was very similar. Significant correlations were also found between the Third Edition and the verbal or English sections in Preliminary Scholastic Assessment Test (PSAT), Scholastic Assessment Tests (SAT), American College Testing Program (ACT), and grade point average (GPAs) (Ekwall & Shanker, 1988; Lipson & Wixson,1991; MacGinitie, MacGinitie, Maria & Dreyer, 2002).

In 2008, Rowe, Ozuru, O'Reilly and McNamara examined the difficulty in different standardized reading tests currently used in the United States. They cross-examined the Level 7/9 and Level 10/12 of the GMRT and concluded that the GMRT contains a variety of passages with varying ranges of difficulty, differing in a number of dimensions. Also, the tests contain questions of several different types;

most of them cannot be answered by simply eliminating distractors. The test extensively measures many different subcomponents implicit in the reading comprehension of the text in the context of various reading circumstances.

Relationships between Reading and Writing

In both of the GSAT and the AST English exams, the subtests are similar with multiple-choice questions (72%), a translation task (8%), and a guided short essay writing (minimum length - 120 words) (20%). In the current study, the GMRT offers no writing tests; yet, many correlational and experimental studies since the 1930s have documented the reading and writing relationships. Loban (1963), in a longitudinal study of students' reading and writing development across 4th, 6th, and 9th grades, indicated strong relationships between reading and writing as measured by test scores. He reported that students who wrote well also read well, and that the converse was also true. Stotsky (1983) published a review of studies that span approximately fifty years from the beginning of the 1930s to 1981. She concluded that "better writers tend to be better readers, that better writers tend to read more than poorer writers, and that better readers tend to produce more syntactically mature writing than poorer readers" (p. 636). According to Smith (1983), reading like a writer allows one to actually become a writer. When reading like a writer, the reader takes in and learns from the author's style, use of conventions and the like. When reading like a writer, the reader uses the author's text as a model for the texts that he or she will eventually write.

METHOD

Participants

The subjects consisted of 224 freshmen at a medical university in central Taiwan. These students were from the five freshman English classes (a total of 242 students) taught by the researcher in the first semester. Freshman English at the university is compulsory, but the students are free to choose the teacher or the time slot most suitable to their class schedule. These participants were coded into two groups: GSAT group and AST group. The AST group consisted of 53 students, who came into the university through the AST Exam held in early July, 2012. The GSAT group comprised 171 students, who came into the university through the GSAT Exam held in January, 2012. The selection of the students for the study excluded overseas Chinese students, foreign students, and the students who came into the university through neither the AST Exam nor the GSAT Exam. Table 1 displays the English performance on respective university entrance examination.

Table 1

Distributions of English Performance on AST and GSAT Exams

AST Group	Mean Score: 68.25						
Score	88.5~80	79~70	69~60	59~50	49~40	39~30	--
N = 53	07	23	13	04	05	01	--
GSAT Group	Mean Rank: 12.21						
Rank	15	14	13	12	11	10	09
N = 171	05	39	40	33	25	16	13

Instrumentation

In the current study, Level 7/9, Form S of the GMRT was used as the measure. The reliability coefficients ranged from 0.94 at Grade 7 to 0.95 at Grade 9 (Cheng, 2010a; MacGinitie, MacGinitie, Maria, & Dreyer, 2002). The 7/9 grade level equating studies showed that the correlations were 0.90 for the 7th graders who took both Level 6 and Level 7/9, 0.90 for the 8th graders who took both Level 6 and Level 7/9, 0.90 for the 9th graders who took both Level 6 and Level 7/9, and 0.87 for the 10th graders who took both Level 7/9 and Level 10/12.

Level 7/9 appeared appropriate for this study as the maturity and difficulty of the content made it an appropriate test for students from Grades 5 to Grade 12.9 (Cheng, 2010a; MacGinitie, MacGinitie, Maria, Dreyer & Hughes, 2007). For example, if a sixth grader takes the Level 6 test, the grade equivalents that can be considered meaningful on that test are from Grade 4.0 to Grade 9.9. That is because the maturity and difficulty of test content are reasonably similar in Level 6 and Levels 4, 5, and 7/9. If seventh and eighth graders take the Level 7/9 test, the grade equivalents considered meaningful on that test are from Grade 5.0 to Grade 12.9. That is because the maturity and difficulty of test content are reasonably similar in Level 7/9 to the two levels below, Levels 5 and 6, and the level above, Level 10/12.

In this study, the GMRT was administered to the students in the regular school class hours in the first week of the first semester. Before the tests, students were intimated about the purpose of the tests and that the scores would not be incorporated in the final grade report. They were encouraged to try their best with a choice to escape the tests if they did not feel comfortable. The answers were manually evaluated by two teachers. Each question was accredited one point, with a total of 93 points for the test.

RESULTS AND DICUSSION

AST Exam English Test and the GMRT Score

Table 2 displays the descriptive data of the AST Exam group. The mean AST Exam English score is 68.245 (Min = 34, Max = 88.5, SD = 12.691). The mean

GMRT raw score is 44.226 (Min = 22, Max = 70, SD = 10.852).

Table 2

AST Group - AST English Score × GMRT Score

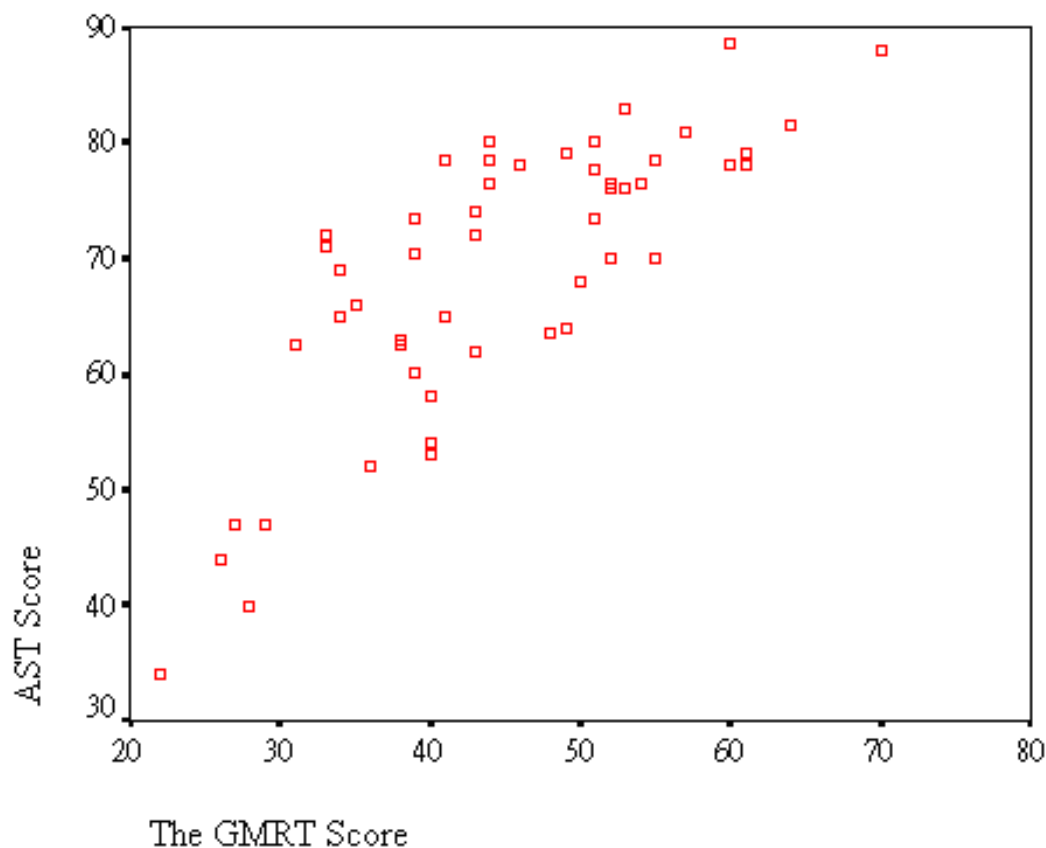|	N	Mean	Min	Max	SD
AST English Score	53	68.245	34.00	88.50	12.691
GMRT Score	53	44.226	22.00	70.00	10.852

The Pearson product-moment correlation coefficient was computed to assess the relationship between the AST Exam English score and the GMRT score. There was a positive correlation between the two variables, $r = 0.801$, n = 53, $p = 0.000$. Table 3 summarizes the result. A scatter plot also illustrates the result (Figure 1). Overall, there was a strong, positive correlation between the AST Exam English test and the GMRT.

Table 3

Correlation Coefficient - AST English Score × GMRT Score

Pearson r	N	Correlation	Sig.
AST English × GMRT Score	53	.801(**)	.000

** Correlation is significant at the 0.01 level.

Figure 1

Scatter Plot

8

GSAT Exam English Rank Score and the GMRT Score

GSAT ranks English score from 1 to 15. Table 4 displays the descriptive data of the GSAT Exam group. The mean GSAT Exam English rank score is 12.216 (Min = 09, Max = 15, SD = 01.606). The mean GMRT raw score is 41.585 (Min = 23, Max = 75, SD = 09.686).

Table 4

GSAT Group - GSAT English Rank Score × GMRT Score

|	N	Mean	Min	Max	SD
GSAT Exam Rank	171	12.216	09.00	15.00	01.606
GMRT Score	171	41.585	23.00	75.00	09.686

The Pearson product-moment correlation coefficient was computed to assess the relationship between the GSAT Exam English score and the GMRT score. There was a positive correlation between the two variables, $r = 0.637$, n = 171, $p = 0.000$. The results are given in Table 5 and Figure 2. Overall, there was a significant and positive correlation between the GSAT Exam English test and the GMRT.

Table 5

Correlation Coefficient - GSAT English Rank Score × GMRT Score

Pearson r	N	Correlation	Sig.
GSAT Exam Rank × GMRT Score	171	.637(**)	.000

** Correlation is significant at the 0.01 level (two-tailed).

Figure 2

Scatter Plot

9
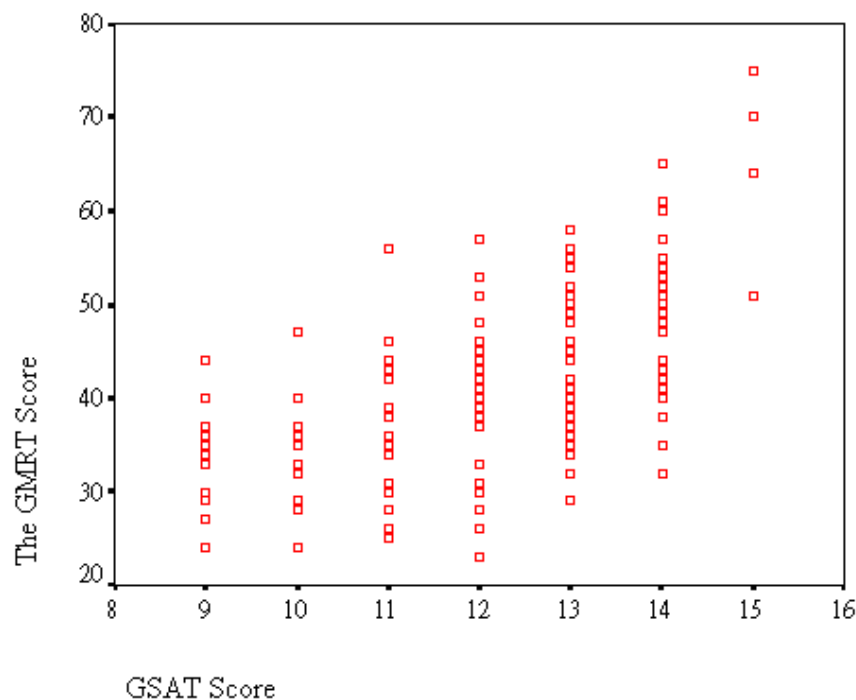
Equating the AST Score with Grade Equivalents

The GMRT raw scores of the students in the AST group were converted to the grade equivalents. Table 6 displays the frequency data. Then, each GE was matched to the minimum AST score. Table 7 displays the results. An ascending trend was observed. It shows that to score at or above Grade 6, a minimum AST English score of 62 is needed. To score at or above Grade 7, a minimum score of 70 is necessary. To score at or above Grade 8, the score is 78; at or above Grade 9, 81.5; and at or above Grade 11, 88. No students scored at Grade 10, 12 and PHS.

The correlation coefficient of 0.801 between the AST and the GMRT indicates $(.801)^2$ or 64.16% common variance. According to Gay (1987), coefficients in the range of .60s to .70s are usually considered adequate for group prediction purposes, and coefficients in the .80s and above for individual prediction purposes. Yet, it is necessary to note that a perfect correlation is quite improbable for any correlational research; therefore, for example, students who scored 88 at the AST English test might not necessarily score at Grade 11. They might score higher or lower. However, to reach Grade 11, based on the AST Exam held in July 2012, an AST English score of 88 is the minimum requirement.

Table 6
Grade Equivalents: AST Group

Grade Level	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 3.00	1	.6	1.9	1.9
4.00	6	3.5	11.3	13.2
5.00	17	9.9	32.1	45.3
6.00	11	6.4	20.8	66.0
7.00	11	6.4	20.8	86.8
8.00	5	2.9	9.4	96.2
9.00	1	.6	1.9	98.1
11.00	1	.6	1.9	100.0
Total	53	100.0		

Table 7
AST Group – GMRT GE × Minimum AST English Score

Grade Equivalent	6	7	8	9	10	11	12	PHS
Min. AST English Score	62	70	78	81.5	--	88	--	--

Equating the GSAT English Scores with Grade Equivalents

The GMRT raw scores of the students in the GSAT group were converted to the grade equivalents. Table 8 displays the frequency data. Then, each GE was matched to

the minimum GSAT ranking score (See Table 9). The GE percent distribution of R15 could not be tallied due to low number of students (5 students only); hence, alternatively a numeral distribution was used. We found one student scored at PHS, one at Grade 11, two at Grade 9, and one at Grade 7. The GE percentage distributions from R14 to R9 showed a descending trend. At R14, the percentage showed that the chance of scoring at or above Grade 6 was 71.79%. The percentage descended to 35.89% at Grade 7; 12.82% at Grade 8; and 02.6% at Grade 9. At R13, the percentage descended from Grade 6, 50%, to 20% at Grade 7, to 2.5% at Grade 8. At R12, the percentage descended from Grade 6, 36.36%, to 9.09% at Grade 7, to 3% at Grade 8. At R11, the percentage descended from Grade 6, 20% to 4% at Grade 7. Yet, at R10 and R9, the chance to score at or above Grade 6 is low.

As the common variance is only 40.58%, it is not possible to predict precisely the grade level at which a student who scores a rank of 14 can reach. Table 9 offers only a rough estimation. Nevertheless, it is more credible to express that 80% of the students who score a rank lower than 12 might have a reading GE lower than Grade 6. Moreover, the number distribution of R15 shows that R15 contains: 1) students who might possess extreme English ability such as PHS level and 2) a significant number of students who might score at or above Grade 9.

Table 8
Grade Equivalents: GSAT Group

Grade Level	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 4.00	28	16.2	16.4	16.4
5.00	71	41.0	41.5	57.9
6.00	41	23.7	24.0	81.9
7.00	20	11.6	11.7	93.6
8.00	6	3.5	3.5	97.1
9.00	3	1.7	1.8	98.8
11.00	1	.6	.6	99.4
13.00	1	.6	.6	100.0
Total	171	100.0		

Table 9
GSAT Group – Percent of GE Distribution × GSAT English Rank Score

Possible %	R14	R13	R12	R11	R10	R09
≧GE 9	02.60	00.00	00.00	--	--	--
≧GE 8	12.82	02.50	03.00	00.00	--	--
≧GE 7	35.89	20.00	09.09	04.00	00.00	00.00
≧GE 6	71.79	50.00	36.36	20.00	06.25	07.69

CONCLUSION

The study shows that English exams developed in Taiwan correlate highly with a popular American standardized norm-referenced reading test. The GMRT also demonstrates that a test not only sorts students but also helps in planning instructional emphases, in locating students who are ready to work with more advanced materials, and in deciding the levels of instructional materials suitable for the students. As the AST and GSAT English tests offer only a score or a rank, most universities in Taiwan tend to retest their students for ability grouping without examining their values in facilitating freshman English teaching. As tremendous time and efforts are devoted to the preparation for the AST and GSAT English tests by the College Entrance Examination Center, high schools, parents, and high school students, their values should be evaluated and justified through other instrumentations such as the GMRT.

A good reading test should offer information that helps teachers locate their students' reading grade levels. Learning will become meaningful if a student's reading ability matches the readability of the textbook s/he is assigned to read (Ekwall & Shanker, 1988; .Lipson & Wixson, 1991). As most EFL teachers in Taiwan are majors in English linguistics, TEFL, and English literature, the tests they use to identify their students' English levels are Test of English as a Foreign Language (TOEFL), the International English Language Testing System (IELTS), and General English Proficiency Test (GEPT). They have insufficient knowledge about the functions of the American-developed reading tests for American graded schools. They should be informed of the functions of other English tests and how they can be applied to enhance their teaching and their students' learning (Cheng, 2010a).

REFERENCES

Burns, P. C. & Roe, B. D. (1989). *Burns/Roe Informal Reading Inventory – Preprimer to Twelfth Grade.* Boston: Houghton Mifflin

Carpenter, R. D. & Paris, S. G. (2005). Issues of validity and reliability in early reading assessments. In Paris, S. G. & Stahl, S. A. (Eds.), *Children's Reading Comprehension and Assessment,* (pp.279~304). Mahwah, NJ: Lawrence Erlbaum Associates.

CEEC. (2013). *GSAT and AST*. http://www.ceec.edu.tw/CeecEnglishWeb/E07 Process_GSAT.aspx and http://www.ceec.edu.tw/CeecEnglishWeb/E07 Process AST.aspx

Cheng, T. Y. (2009). General English Vocabulary Levels among College High-achievers in Taiwan. *Journal of Applied Foreign Languages*, 12, 49~78.

Cheng, T. Y. (2010a). Readiness of College Students in Taiwan to Read to Learn from Texts in English. *Asian EFL Journal-Professional Teaching Articles*, 41, 24~49.

Cheng, T. Y. (2010b). Taiwanese College Teachers' Attitudes towards English Reading Instructions in Discipline-specific Areas. *Asian Journal of English Language Teaching (AJELT)*, 20, 117~134.

Cook, B.G., Gerber, M.M., & Semmel, M.I. (1997). Are effective schools reforms effective for all students? - The implications of joint outcome production for school reform. *Exceptionality,* 7 (2), 77~96.

Drummond, K., Chinen, M., Duncan, T. G., Miller, H. R., Fryer, L., Zmach, C. & Culp, K. (2011). *Impact of the Thinking Reader software program on grade 6 reading vocabulary, comprehension, strategies, and motivation* (NCEE 2010-4035). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Ekwall, E. E. & Shanker, J. L. (1988). *Diagnosis and remediation of the disabled reader*. Needham Heights, Massachusetts: Allyn and Bacon.

Fisher, G. (2001). Cross age tutoring: Alternatives to the reading resource room for struggling adolescent readers. *Journal of Instructional Psychology,* 28 (4), 234~240.

Gay, L. R. (1987). *Educational research: Competencies for analysis and application* (3rd Edition). Columbus, OH: Merrill Publishing.

Gilbert, S. L. (2009). *An examination of the effects of computer assisted instruction on reading comprehension*. UMI Microfilm No. 1465864. ProQuest LLC.

Johnson, K.M. & McCabe, P.P. (2003). Review of the Gates–MacGinitie 4th edition, forms S and T. In *Fifteenth mental measurements yearbook*. Lincoln, NE: Buros Institute Inc.

Loban, W. (1963). T*he Language of Elementary School Children*. Urbana IL: National Council of Teachers of English.

Lipson, M.Y. & Wixson, K.K. (1991). *Assessment & instruction of reading disability: An interactive approach*. New York: HarperCollins.

LTTC (the Language Training & Testing Center). (2003). *Concurrent validity studies among the GEPT, Ability Exam English score, CBT TOEFL and CET-6*. http://www.lttc.ntu.edu.tw/research/statistics/researchreport.pdf

MacGinitie, W.H., MacGinitie, R.K., Maria, K. & Dreyer, L.G. (2000). *Gates-MacGinitie Reading Tests –Directions for Administration (levels 7/9 & 10/12)* (4th Ed.). Rolling Meadows, IL: Riverside.

MacGinitie, W.H., MacGinitie, R.K., Maria, K. & Dreyer, L.G. (2002). *Gates-MacGinitie reading tests – Technical report (Forms S and T)* (4th Ed.). Rolling Meadows, IL: Riverside.

MacGinitie, W.H., MacGinitie, R.K., Maria, K., Dreyer, L.G. & Hughes, K.E. (2007). *Gates-MacGinitie Reading Tests – Manual for scoring and interpretation*

(levels 7/9 & 10/12) (4th Ed.). Rolling Meadows, IL: Riverside.

Nelson, R.N. & Stage, S.A. (2007). Fostering the development of vocabulary knowledge and reading comprehension though contextually-based multiple meaning vocabulary instruction. *Education & Treatment of Children,* 30 (1), 1~22.

Paris, S.G. & Associates. (2004). Assessing the effectiveness of summer reading programs. In Borman, G.D., & Boulay, M. (Eds.), *Summer Learning: Research, Policies, and Programs,* (pp. 121-162). Mahwah, NJ: Lawrence Erlbaum.

Rowe, M., Ozuru, Y., O'Reilly, T. & McNamara, D. S. (2008). Where's the difficulty in standardized reading tests: The passage or the question? *Behavior Research Methods*, 40 (4), 1001~1015.

Singer, H. & Donlan, D. (1989). *Reading and learning from text*. Hillsdale, NJ: Lawrence Erlbaum.

Slossen, R. L. (1988). *Slossen Oral Reading Test (SORT)*. New York: Slossen Educational Publications.

Smith, F. (1983). Reading like a writer. *Language Arts*, 60 (5), 58~567.

Stotsky, S. (1983). Research on Reading /Writing Relationships: A Synthesis and Suggested Directions. *Language Arts*, 60(5), 627~642.

Tatum, A.W. (2004). A road map for reading specialists entering schools without exemplary reading programs: Seven quick lessons a reading specialist who worked in a high-poverty; urban elementary school provides seven lessons that can serve as guideposts for others. *The Reading Teacher,* 58 (1), 28~38.

Tilstra, J., McMaster, K., Broek, P. V.D., Kendeou, P. & Rapp, D. (2009). Simple but complex: components of the simple view of reading across grade levels. *Journal of Research in Reading*, 32 (4), 383~401.

Walt, J. L. & Steyn, F. Jr. (2008). The validation of language tests. *Stellenbosch Papers in Linguistics*, 38, 191~204