

# Using Computer Aided Assessment System to Assess College Students Writing Skill

Huihua He

College of Education, Shanghai Normal University, Shanghai, China  
Email: hehuihua@shnu.edu.cn

Jin Liu and Hongmin Ren

College of Information Engineering, Shanghai Maritime University, Shanghai, China  
Email: {jinliu, hmren}@shmtu.edu.cn

**Abstract**— College students are facing challenges to present their ideas by writing a paper because they rely more on information from computer and web. The purpose of this paper is to present a novel computer aided assessment system, to assess college students' writing ability. The CAAS system comprises of an expert team, a set of achievement standard for assessment, and software systems to conduct data analysis and store related information. It has been used by institute in U.S. It yields face validity and consistency reliability. Objective evaluation results will be provided by randomly-assigned multiple reviewers. CAAS can assist college students improve writing skills, can describe a brief picture and detailed information for both school administrators and policy makers as well.

**Index Terms**—Writing Ability; College Student; Assessment System; Database; Web; Computer Software

## I. INTRODUCTION

College educators in all disciplines have been concerned in recent years about the apparent decline in the writing ability of their students. As part of a recent study, faculty members from 190 departments and 34 universities in Canada and the United States complete a questionnaire on the importance of academic writing skills in their fields: business management, psychology, computer science, chemistry, civil engineering, and electrical engineering [1]. In all six areas, writing ability was judged important to success in undergraduate and graduate training. One can assume that it is not merely important to western countries but essential to eastern countries. In all colleges and universities worldwide, even first-year student in programs such as electrical engineering must write laboratory reports and article summaries. Hiring managers from various institutions also agreed that writing ability is even more important to professional than to academic success.

Understandably, there has been a growing interest among educators in methods used to measure, evaluate, and predict college students' writing skills. However,

little is known about the Chinese college students' writing skills, especially the assessment systems. Generally speaking, educators use assessments to hold student accountable for learning. In order to collect reliable and valid information, different methods have been evolved. According to literatures from western countries, there are two types of assessments regarding college students' writing ability: direct assessment and indirect assessment. "Direct assessment" requires the examinee to write an essay or several essays, typically on preselected topics.

Indirect assessment usually requires the examinee to answer multiple-choice items. Hence direct assessment is sometimes referred to as a "production" measure and indirect assessment as a "recognition" measure. [2] Richard Stiggins [3] observes that indirect assessment tends to cover highly explicit constructs in which there are definite right and wrong responses (e.g., a particular language construction is either correct or it is not). Direct assessment, on the other hand, tends to measure less tangible skills (e.g., persuasiveness), for which the concept of right and wrong is less relevant.

Each method has its own advantages and disadvantages. The debate between these two approaches were drawn chiefly over the issues of validity (the extent to which a test measures what it purports to measure—in this case, writing ability) and reliability (the extent to which a test measures whatever it does measure consistently). For example, Orville Palmer [12] wrote that essay tests (indirect assessment) are neither reliable nor valid, but objective tests do constitute a reliable and valid method of ascertaining student compositional ability. In a landmark study of the same year, Godshalk, Swineford, and Coffman [11] showed that, under special circumstances, scores on brief essays could be reliable and also valid in making a unique contribution to the prediction of performance on a stable criterion measure of writing ability.

In addition, the objective tests not only could achieve extremely high statistical reliabilities but also could be administered and scored economically, thus minimizing the cost of testing a growing number of candidates. However, multiple-choice tests do not seem to measure writing ability because the examinees do not write.

---

Sponsored by Shanghai Pujiang Talents Program (No. 10PJC063); Shanghai Normal University Scientific Research Initial Fund (No. PW933); Shanghai Science Foundation for The Excellent Youth Scholars (No. ssd10025).

Along with the popularity of direct assessment has grown the tendency of writing tests, especially in higher education, to focus on "higher-level" skills such as organization, clarity, sense of purpose, and development of ideas rather than on "lower-level" skills such as spelling, mechanics, and usage [4].

Therefore, direct assessment has rapidly gained adherents. One reason is that, while direct and indirect assessments appear statistically to measure very similar skills, indirect measures lack "face validity" and credibility among English teachers. That is, multiple-choice tests do not seem to measure writing ability because the examinees do not write. Also, reliance on indirect methods exclusively can entail undesirable side effects. Students may presume that writing is not important, or teachers that writing can be taught -- if it need be taught at all -- through multiple-choice exercises instead of practice in writing itself. One has to reproduce a given kind of behavior in order to practice or perfect it, but not necessarily in order to measure it. Still, that point can easily be missed, and the College Board has committed itself to supporting the teaching of composition by reinstating the essay in the Admissions Testing Program English Composition Test.

Earle G. Eley (1955)[10] has argued that "an adequate essay test of writing is valid by definition since it requires the candidate to perform the actual behavior which is being measured" (p. 11). In direct assessment, examinees spend their time planning, writing, and perhaps revising an essay. But in indirect assessment, examinees spend their time reading items, evaluating options, and selecting responses. Consequently, multiple-choice tests confound the variables by measuring reading as well as writing skills, or rather editorial and error-recognition skills, some insist, in that the examinees produce no writing.

Multiple-choice tests, of course, do not require -- or even permit -- any such performance. And they have also been criticized for making "little or no attempt to measure the 'larger elements' of composition, even indirectly" (Braddock, Lloyd-Jones, & Schoer, 1963, p. 42).[8] Their critics generally maintain that objective tests cannot assess originality, creativity, logical coherence, effectiveness of rhetorical strategy, management and flexibility of tone, the ability to generate ideas and supporting examples, the ability to compose for different purposes and audiences, the ability to stick to a subject, or the ability to exercise any other higher-level writing skill -- in short, that objective tests cannot assess anything very important about writing.

Although objective tests do not show whether a student has developed higher-order skills, much evidence suggests the usefulness of a well-crafted multiple-choice test for placement and prediction of performance. Moreover, objective tests may focus more sharply on the particular aspects of writing skill that are at issue. Students will pass or fail the "single writing sample quickly scored" for a variety of reasons -- some for being weak in spelling or mechanics, some for being weak in grammar and usage, some for lacking thesis development and organization, some for not being inventive on

unfamiliar and uninspiring topics, some for writing papers that were read late in the day, and so on. An essay examination of the sort referred to by Cooper and Odell [9] may do a fair job in ranking students by the overall merit of their compositions but fail to distinguish between those who are grammatically or idiomatically competent and those who are not.

Both essay and objective tests entail problems when used alone. Would some combination of direct and indirect measures then be desirable? From a psychometric point of view, it is sensible to provide both direct and indirect measures only if the correlation between them is modest--that is, if each exercise measures something distinct and significant. The various research studies reviewed by Stiggins (1981)[3] "suggest that the two approaches assess at least some of the same performance factors, while at the same time each deals with some unique aspects of writing skill...each provides a slightly different kind of information regarding a student's ability to use standard written English" (pp. 1,2). Godshalk et al. (1966)[22] unambiguously state, "The most efficient predictor of a reliable direct measure of writing ability is one which includes essay questions...in combination with objective questions" (p. 41). The substitution of a field-trial essay read once for a multiple-choice subtest in the one-hour ECT resulted in a higher multiple correlation coefficient in six out of eight cases; when the essay score was based on two readings, all eight coefficients were higher. The average difference was .022--small but statistically significant (pp. 36-37). Not all multiple-choice subtests were developed equal, though. In every case, the usage and sentence correction sections each contributed more to the multiple prediction than did essay scores based on one or two readings, and in most cases each contributed more than scores based on three or four readings (pp. 83-84).

The reliability of the direct measure can also be enhanced by combining the essay score and the objective score in one total score for reporting. Several College Board tests employ this approach. And "assessments like the ECT combine direct and indirect assessment information because the direct assessment information is not considered reliable enough for reporting. A single ECT score is reported by weighting the direct and indirect information in accordance with the amount of testing time associated with each. Such a weighting may not be the most appropriate," Breland and Jones (1982) [7] maintain, "Another approach to combining scores might be to weight scores by relative reliabilities" (or, perhaps, validities). They add that "reliable diagnostic sub-scores could possibly be generated by combining part scores from indirect assessments with analytic scores from direct assessments" (p. 28).

Combined measures seem to be a reasonable and acceptable approach to evaluate college students' writing skills. However, there is a little difficult for Chinese college students. First, Chinese college students seem to be more care about their English than their first language. They spend more time on all kinds of English Tests than Chinese writing. Second, the population of Chinese

undergraduate students seems to be larger and larger. The ratio between undergraduates and faculty members is getting lower and lower. So there are not enough advisors to read and grade students' writings. Last but not the least, there is lacking an achievement standard system for Chinese college students writing ability.

In this study, a computer aided assessment system will be established for students, educators and legislators. The rest of this paper will provide detailed description of the whole system, the reliability and validity of this method and future concerns on this assessment system.

## II. COMPUTER-AIDED WRITING ABILITY ASSESSMENT SYSTEM FOR COLLEGE STUDENT

There are four major components in this computer aided writing ability assessment system: students' writing portfolio, expert pool, achievement standards and software system. The detail of each part is brought in following sections, and we will use CAAS as the short for our Computer Aided Assessment System.

### A. Students' Writing Portfolio

CAAS will create a portfolio for each freshman when he/she first registers at student office. Student must submit their writing assignments into the CAAS system as required. The Writing Portfolio consists of two parts:

- The Packet (three samples of writing assignments from any course that student take);
- An Academic Paper (a two-part, impromptu essay).

Total of four entries must be submitted to the assessment system during four years. Generally speaking, student will be encouraged to submit one assignment per year. There is a guideline for students to collect their work. For example, possible categories that writing portfolio package could include:

- Reflective Writing
- Personal Expressive Writing
- Personal Narrative--focusing on one event in the life of the writer
- Memoir
- Literary Writing(s)
- Short story - Poem - Script - Play

An online submission system will guide student to submit their work through internet. Each student will have their student ID to sign in and confirm their information, such as name, age, major and department. Then they can upload their work. If any student only can provide hardcopies, they can ask staff to help them scan the hardcopy and upload to the system. Once an entry is received, it will be randomly assigned to two experts to grade. CAAS System is responsible to pick the reviewer and keep record for every student.

TABLE I.  
SAMPLE EXPERT GRADING FORM

<b>Discourse Characteristics</b>	1	2	3	4	5	6	<b>Paper Digest</b>
Statement of thesis		•					
Overall organization	•						
Rhetorical strategy			•				
Noteworthy ideas		•					
Supporting material						•	
Tone and attitude		•					
Paragraphing&transition					•		
Sentence variety			•				
Sentence logic			•				
<b>Syntactic Characteristics</b>	1	2	3	4	5	6	<b>Paper Digest</b>
Pronoun usage		•					
Subject-verb agreement		•					
Parallel structure			•				
Idiomatic usage				•			
Punctuation			•				
Use of modifiers			•				
<b>Lexical Characteristics</b>	1	2	3	4	5	6	<b>Paper Digest</b>
Level of diction			•				
Range of vocabulary				•			
Precision of diction	•						
Figurative language			•				

### B Expert Team

An expert pool will be established for the assessment system. The pool is composed of experienced university faculty members who volunteer to grade students' papers. For each review session, the expert team members are randomly chosen from the pool by CAAS. Number of team members is determined by how many students take part in the review. Usually each member will be assigned 5 anonymous papers which are randomly picked from all the submitted ones. Moreover, there are two extra rules to make the grade result objective and fair. One is that CAAS will make sure that the chosen reviewers are from different discipline than author of the paper. The other is that each reviewer won't be assigned papers from same student twice. Reviews will follow the standards presented in next part to grade paper, and submit grading results, paper digest and comments to system by the deadline enforced by CAAS.

### C Achievement Standards

There are a few criteria for Assessment used by Writing Portfolio Reviewers:

- Conception of assigned task, so that the reviewers can see how well students comprehend and make decisions about what they write.
- Focus on a main point, so that the reviewers can see a clear relationship between the main point and the topic.
- Organization of ideas, so the reviewers can understand the flow of students' development from beginning to middle to end.
- Support of the points, so the reviewers can see how students' ideas and information are related

to the topic, and is persuaded of the points that are developing.

- Proofreading of the written work, so that errors of grammar, usage, and diction do not seriously distract the reviewer.

Each expert will grade according the characteristics shown in Table 1 which are marked as following levels: 1=very strong, 2=somewhat strong, 3=strong, 4=weak, 5=somewhat weak, 6=very weak. As picking reviewer process and assigning paper process are all randomly determined by CAAS, and the review standard is equally enforced by all the reviewers, there is no bias on which faculty will be assigned a particular topic paper or some different standard will be applied. Thus we make sure review process conducted by the expert team is fair to everyone.

In addition to the coordinated series of writing assignments, we also do the following: (1) we invite our experts to deliver a lecture to students at the beginning of each course in the writing initiative sequence that emphasizes the importance of writing in their future profession, describes our writing initiative, and details the writing requirements and available writing-assistance resources; (2) experts will provide significant feedback on students' writing performance, in terms of both professional content and grammatical contentions; including feedbacks on grammar and style issues while we grade the assignments for content; (3) we provide students with specific writing reference materials, including an Internet web page tailored to our writing initiative.

#### D Software System

The software system for Writing Portfolio is designed to facilitate the whole assessment program. Thus it needs to store students' submissions, store reviewers' comments, distribute papers to reviewers, record all the related information, and conduct a set of analysis work. As shown in Fig. 1, it comprises of a web interface front end, a database back end and software functionality set. All software is based on Windows platform; ASP.NET and SQL server are used to implement the system.

There are mainly three types of users for CAAS, namely students, reviewers and administrators. For students, they use CAAS as an interface to submit their work, check their scores and get up to date information regarding the assessment procedure. For reviewers, CAAS is a tool to help them have idea of current review process, get students' paper assigned to them, save their grading work and comments. For administrators, CAAS provides all the functionalities they need to run the assessment program. The main administrative functionalities are shown as follows:

- Data search. With fully customized intelligent data search, administrator can easily find the record for any students who had attend this program with their name, ID or even some ambiguous information such as their major or year of graduation etc.

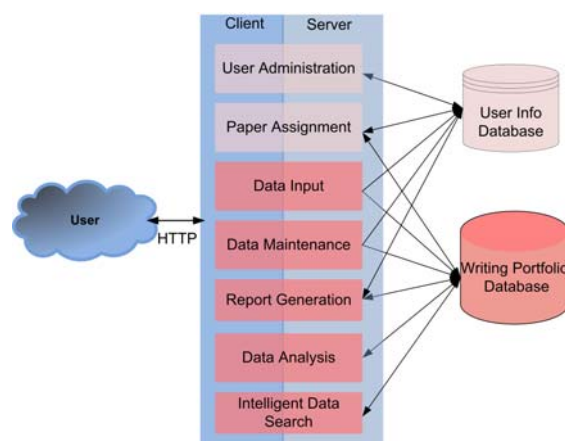


Figure1. CAAS system block diagram.

- Data analysis. CAAS provides a set of analyzing ability which is essential to the assessment program. The detail will be presented in next section.
- Data maintenance. Provide functionalities to maintain data in both User information database and Writing Portfolio database, such as modify, delete, overwrite and backup etc.
- Data input. This module is responsible to take all sorts of input to the system. Students, reviews or administrator can use this module to input their personal information to set up each one's account. In order to let students submit their paper, system provides interface to allow them upload .doc, .pdf or .jpg file.
- Report Generation. Based on the data analysis results, CAAS can automatically generate the customizable report to show the overall status of current assessment program by present all the interested statistics such as comparison of performance of transfer students and non-transfer students, statistics of students in each calendar year, statistics for improvements made by students in different year or major.
- Paper Assignment. This module implements the idea presented in Section B to randomly assign students papers to different reviewers while following the rules to keep assignment fair and objective. The assignment decision will only be available to administrator. Reviews won't know whose work he/she is grading and students won't know who is grading his/her work either.

#### E Data Analysis

The CAAS system can provide a set of data analysis functions, such as descriptive analysis and validation data analysis. Descriptive findings offer insights into the status of student writing performance at the university through the Writing Portfolio, such as "average time to submit the papers", "performance for all students", and "performance according to student's gender, academic level and academic area".

Table II  
SAMPLE REPORT TABLES

Task	Tier I Rating			Final Rating		
	B-E	A	F	B-E	A	F
Resolving	61.3	10.4	28.1	81.5	8.13	10.2
Solving	62.2	9.72	27.9	82.3	8.01	9.66
Analyzing	56.5	10.9	32.4	83.5	4.94	11.5
Choosing	60.1	9.11	30.7	80.9	7.37	11.6

(A) TIER I AND FINAL RATINGS: ALL STUDENTS, 200X-200Y (IN PERCENTAGE)

Topic	200x-200y			200z-200y		
	B-E	A	F	B-E	A	F
24 Freeway building	50	0	50	57.1	14.2	28.5
4 Read vs. Television	48.2	11.0	40.6	60.4	10.4	29.0
19 American higher education shows strong class	54.8	8	37.1	58.4	9.43	32
32 Sports clichés	54.7	9.05	36.2	54.7	9.05	36.2
14 Malls lead to consumerism	55.6	9.15	35.1	59.6	9.80	30.5
22 Immigration of wealthy internationals	60	5.33	34.6	60.0	6.84	33.1
10 American idea of success is mere acquisition of goods	53.3	13.3	33.3	61.6	8.33	30.0
21 America as a warrior nation	55.6	11.2	33.0	60.0	9.40	30.3
7 Taking photographs of private citizens is unethical	58.8	9.60	31.5	61.0	9.50	29.4
15 Television undermines the habit of book reading	55.1	13.7	31.0	64.4	11.3	24.1
20 Racial hate messages on campus	62.7	8.18	29.0	62.6	8.71	28.6
27 Banning offensive language	61.0	10.5	28.4	56.1	11.4	32.4
3 Zoos conceal a human antagonism to animals	59.4	12.3	28.2	59.9	11.0	29.0
30 Web makes research appear easy	63.7	10.1	26.1	61.2	12.9	25.8

(B) TIER I RATINGS, RANKED BY F RATE, ALL STUDENTS, 200X-200X (IN PERCENTAGE)

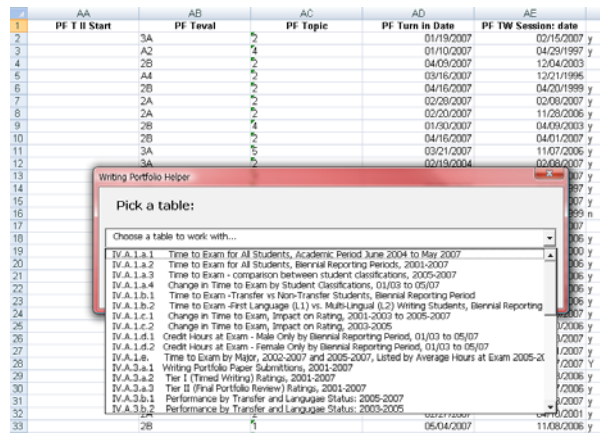
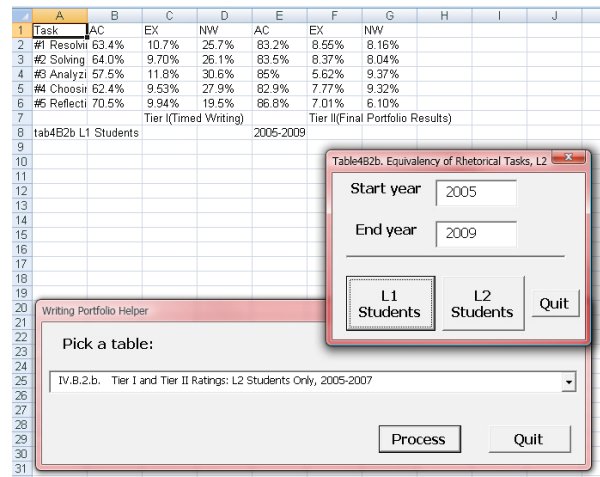


Figure2. Screenshots of data analysis tools

Validation findings provide information that validates the Writing Portfolio as an assessment of undergraduate writing ability. The Writing Portfolio was designed to provide diagnostic feedback regarding the preparedness of undergraduate students to their future jobs, the sample result tables such as “Tier I And Final Ratings”, and “Tier I Ratings , Ranked by Fail Rate”, etc, are shown in Table II.

CAAS employs a VBA [5] program to conduct the analysis work. It can do all the needed statistic work for the program. Here we list some of its functionalities for illustrating purpose as follows:

- Make comparison between student classifications,
- Get change in time to exam by student classifications,
- Get time to exam by major,
- Get performance according to gender,
- Get credit hours at exam by gender,
- Get rating by different tiers,
- Get performance by campus

Figure 2 shows screenshots for the running data analysis tool to get data that administrator is interested in.



### F Results

About a year after we implemented the writing skill assessment system, we began an outcome assessment of its effectiveness in terms of students' writing skill improvements. Using a between-subjects experimental design, we compared the writing skills of 100 students who participated in our writing initiative with the writing skills of 98 students who did not participate in our writing initiative. Using a within-subjects experimental design, we tracked improvements of 60 students' writing skills over time. Results from both outcome assessment methodologies indicate that students' writing skills improve as students participate in our writing initiative. From a curriculum design perspective, our results suggest the need for both general writing experiences, and professionally meaningful writing experiences in the development of skilled adult writing. We find that general writing experiences (e.g., writing in college English classes), regardless of context, are still important as late as the sophomore year of college. However, as students advance to a stage in their studies at which they must begin demonstrating the writing skills that they will need "on the job," professionally meaningful writing assignments are vital to continued development of writing skills.

Based on the results we obtained, computer-assisted evaluation appears to be feasible and produces useful results. These results support the recommendation of the National Commission on Writing in America's School and Colleges that technology be applied to the evaluation of writing to overcome time and effort factors that inhibit the teaching of writing.

### G Future work

CAAS can give objective assessment result which helps students to improve their writing skills and lets school administration gain the idea of the status of students' writing skill. But since this system was only tested in U.S. institute, many work need to be done to make it adapt to Chinese institutes and be useful in Chinese writing ability assessment. We are modifying the criteria for assessment to make it conform to Chinese writing skills.

A student's ability to think critically regarding the rhetorical specifications of an assignment—the way the student addresses the purpose, audience, and subject of an assignment—can best be assessed through multiple observations of student writing contained in portfolios. Hence, in increasing our accuracy with further study, we can increase the defensibility of this methodology. In addition, in that the prohibitive, time-consuming cost of portfolio scoring is well known, the relationship between portfolio scores may identify ways to use cost-efficient, computer-based scores of surface features in conjunction with human-based portfolio scores of rhetorical ability. If such assessment information is then used to enhance instruction, the distance between assessment and instruction may be more readily bridged through an increase in the use of technology.

Moreover, there are plenty of modifications need to be done in the software component of CAAS. To make it use in Chinese institute, the translation work must be done. Also, to avoid copyright problem using proprietary software, we're planning transport the system to LAMP [6] platform. For historical reason, the analysis tools currently are not implemented seamlessly with CAAS database system; we're merging them to make it more convenient for administrators to use.

### III. CONCLUSION REMARK

We have presented a novel computer aided assessment system to assess the writing skill of college students. The system yields enough face validity and can ensure consistency reliability by having multiple reviewers. In addition, it can provide students with objective assessment result, which can help students to improve their writing skills. Computer aided system is applicable, effective and convenient. School administrators can get a big picture and detailed information on students' writing ability. It is also useful for policy makers to provide professional helps or recommendations for college students to ensure their success in the future.

### ACKNOWLEDGMENT

This work was supported by Shanghai Pujiang Talents Program (No. 10PJC063); Shanghai Normal University Scientific Research Initial Fund (No.PW933); Shanghai Science Foundation for The Excellent Youth Scholars (No.ssd10025).

### REFERENCES

- [1] Bridgeman, B., & Carlson, S. Survey of academic writing tasks required of graduate and undergraduate foreign students (TOEFL Research Report No. 15). Princeton, NJ: Educational Testing Service, 1983.
- [2] Breland, H.M., Jones, R.J. Perceptions of writing skill. New York: College Entrance Examination Board, 1982.
- [3] Stiggins, R.J. An analysis of direct and indirect writing assessment procedures. Portland, OR: Clearinghouse for Applied Performance Testing, Northwest Regional Educational Laboratory, 1981.
- [4] Cooper, P. L. The Assessment of Writing Ability: A Review of Research. GRE Board Research Report GREB No. 82-15R. Educational Testing Service, Princeton, NJ, 1984.
- [5] <http://msdn.microsoft.com/en-us/office/aa905411.aspx>
- [6] [http://en.wikipedia.org/wiki/LAMP\\_%28software\\_bundle%29](http://en.wikipedia.org/wiki/LAMP_%28software_bundle%29)
- [7] Breland, H.M., Jones, R.J. (1982). Perceptions of writing skill. New York: College Entrance Examination Board.
- [8] Braddock, R., Lloyd-Jones, R., & Schoer, L. (1963). Research in written composition. Urbana, IL: National Council of Teachers of English.
- [9] Cooper, C.R., & Odell, L. (Eds.). (1977). Evaluating writing: Describing, measuring, judging. Urbana, IL: National Council of Teachers of English.
- [10] Eley, E.G. (1955). Should the General Composition Test be continued The test satisfies an educational need. College Board Review, 25, 9-13.

- [11] Godshalk, F. I., Swineford, F., 6 Coffman, W. E. (1966). The measurement of writing ability. New York: College Entrance Examination Board.
- [12] Palmer, (1966). Sense or nonsense The objective testing of English composition. In C.I. Chase & H.G. Ludlow, (Eds.). Readings in Educational and Psychological Measurement (pp. 284-291). Palo Alto: Houghton.

**Huihua He** received her B.A. degree from East China Normal University, Shanghai, China, her M.A. degree of Human Development and her Ph.D degree of Educational Psychology from Washington State University, Pullman, WA, USA, and conducted her post doctoral research at Washington State University.

She is currently an assistant professor at the College of Education, Shanghai Normal University, Shanghai, China. Her research interests include Research Methodology, Education Assessment and Evaluation, Early Childhood Educaiton.

**Jin Liu** received the B.S. degree from Lanzhou University, Lanzhou, China, the M.S. degree from University of Electrical Technology of China, Chengdu, China, and the Ph.D degree from Washington State University, Pullman, WA, USA. He is currently assistant professor at the College of Information

Technology, Shanghai Maritime University, Shanghai, China. He had held industrial position at Windows Core Networking division, Microsoft Corporation, WA, USA.

His research interests include Network on Chip, Internet Protocols, Wireless Network, software architecture. He is a member of IEEE.

**Hongmin Ren** received the M.S. degree from Dalian University of Technology, Dalian, China, the Ph.D degree from Fudan University, Shanghai, and conducted his post doctoral research at Fudan University and Changjiang Computer Group. He is currently associate professor at the College of Information Technology, Shanghai Maritime University, Shanghai, China. His research interests include software architecture, software components technology and formalization methods.

Dr. Hongmin Ren had co-authored over 20 journal and conference papers, managed 5 provincial ministerial level and bureau level research projects, and taken part in three 863 research projects and Shanghai key technology research project.