



Received: 02 October 2016  
Accepted: 30 January 2017  
Published: 22 February 2017

\*Corresponding author: Tamim Ahmed Khan, Department of Software Engineering, Bahria University, Islamabad, Pakistan  
E-mail: [tamim@bui.edu.pk](mailto:tamim@bui.edu.pk)

Reviewing editor:  
Yvonne Xian-han Huang, Hong Kong Institute of Education, Hong Kong

Additional information is available at the end of the article

## STUDENT LEARNING, CHILDHOOD & VOICES | RESEARCH ARTICLE

# Development of online RAW achievement battery test for primary level

Maria Hanif<sup>1</sup>, Tamim Ahmed Khan<sup>1\*</sup>, Uzma Masroor<sup>2</sup> and Amira Amjad<sup>2</sup>

**Abstract:** Achievement test is a mechanism to measure student's knowledge and abilities. Numerous categories of achievement tests have been developed by different scholars and psychologists. Since they do not directly consider curriculum adopted during the course of study of students, they do not reflect truly upon the achievements of students. We propose an achievement test which is computerized and is based on assessment of RAW (reading, arithmetic and writing) capabilities considering curriculum used for imparting education. We set compositions and contents according to age group and educational standards. We then conduct a series of experiments to show how an achievement test linked with a curriculum is reflective, in a better manner, of the student's achievement index than a general one. We call Online RAW Achievement Battery test and we also develop an application which use for conducting our experiment and formulation of results. Finally, we analyze our results with students' historical records and WRAT-4 which is a well-known standardized test and report our findings.

**Subjects:** Health Psychology; Philosophy of Mind; Testing, Measurement and Assessment

**Keywords:** achievement test; primary level; reading test; arithmetic test; writing test; online RAW

### 1. Introduction

Tests are familiarized to judge student's learning capability, performance, and academic level. Tests fall in different categories e.g. Intelligence tests, personality tests, verbal, non-verbal performance tests, aptitude tests, and achievement test. Each testing system needs some kind of general arrangement in order to assess and evaluate student's performance level and also learning disorder through standardized technique. While developing such tests, we define what precisely is to be measured? What types of score translations are required? What test design or blend of arrangements is required for a valid evaluation? What methodology will be utilized i.e. paper-pencil or computerized based (American Educational Research Association, 1999). Achievement tests play an important role in evaluation of learning skills at any level of training.

### ABOUT THE AUTHORS

Tamim Ahmed Khan did his bachelor of software engineering from United Kingdom (Sheffield University UK), MBA (Finance and Accounting) and MS in Computer Engineering from Pakistan. He obtained Doctorate of Philosophy in Software Engineering from Leicester University, UK.

He served at various positions both in industry and educational institutions. He published several research publications in leading international conferences and research journals and has co-authored a book as well.

### PUBLIC INTEREST STATEMENT

We propose an achievement test which is computerized and is based on assessment of reading, arithmetic and writing capabilities considering curriculum used for imparting education. We also develop a computer application so that educational psychologist and therapist can easily calculate performance score of students efficiently through computerized way to judge students learning disorder. Our work is significant in the sense that the test is automated and our main focus is to design this achievement for specific age group.

We can measure students' knowledge, performance, and abilities in a standardized and systematic way through achievement tests (Gay, 1996). We can also improve standard of education through them as they provide proper feedback about student performance. However, it is pertinent to mention that most of these tests are based on western education system and do not clearly reflect upon students' achievements in the developing countries.

Standardized achievement tests play a critical role in providing an objective feedback to educators in order to judge how much students learned and understand. Educational institution use assessment to judge learner's performance level. Achievement test provides a snapshot of student's performance, standardized tests serve as a tool to decide educational resources and deliver helpful technique to measure learner progress. We propose an approach to evaluate students' performance considering curriculum taught to effectively help clinicians to judge possible learning disorders. In order to develop achievement test, we select textbooks for selection of questions/items and stores in databank. We work in close collaboration with psychologists to finalize on length of our test, preparation of items and outcomes. In order to validate our proposal, we conduct expertise judgment and pre-testing which we call Test Round No. 1 to validate items. We finally run our system and a well-known system which we call Test Run No. 2 and reflect on the outcomes while concluding our results.

The rest of the paper is organized as follows; the methodology is described in Section 2, implementation is presented in Section 3. Section 3.2 presents validation and evaluation furthermore Section 3.3 of this paper presents "students' performance using RAW" and in last we have presented evaluation of Online RAW in Section 3.5.

## 2. Our methodology

We present our methodology for the development of achievement test as explained in Figure 1. In order to develop an achievement test, we first conduct data collection in which we select questions that are fundamental units of tests. We collect data from textbooks of primary level students and we call these questions as "items." An item could be a multiple-choice question, a true/false, short-answer, etc. and we select test items considering books from local as well as international origin (the content of items was taken from commonly taught material which was uniform among all levels of school systems, which was also compatible with international standards of age appropriate level of learning). While selecting items for our achievement test, we consider two subjects i.e. English and Mathematics of primary level and cover two local textbooks and one international curriculum. The local textbooks are approved by the local education boards where the International curriculum is based on Maria Montessori.

Once the books for content is selected, we do content analysis to see if the selection is presenting is truly representative of the complete curriculum and is strong enough to meaningfully judge the ability of their subjects i.e. students (The committee approach was called comprising educational experts i.e. teachers). The finalized selection of contents from selected text books contains the following mixture:

### I. Reading test

- 1.1.1. English Letter Reading Test
- 1.1.2. English Word Reading Test

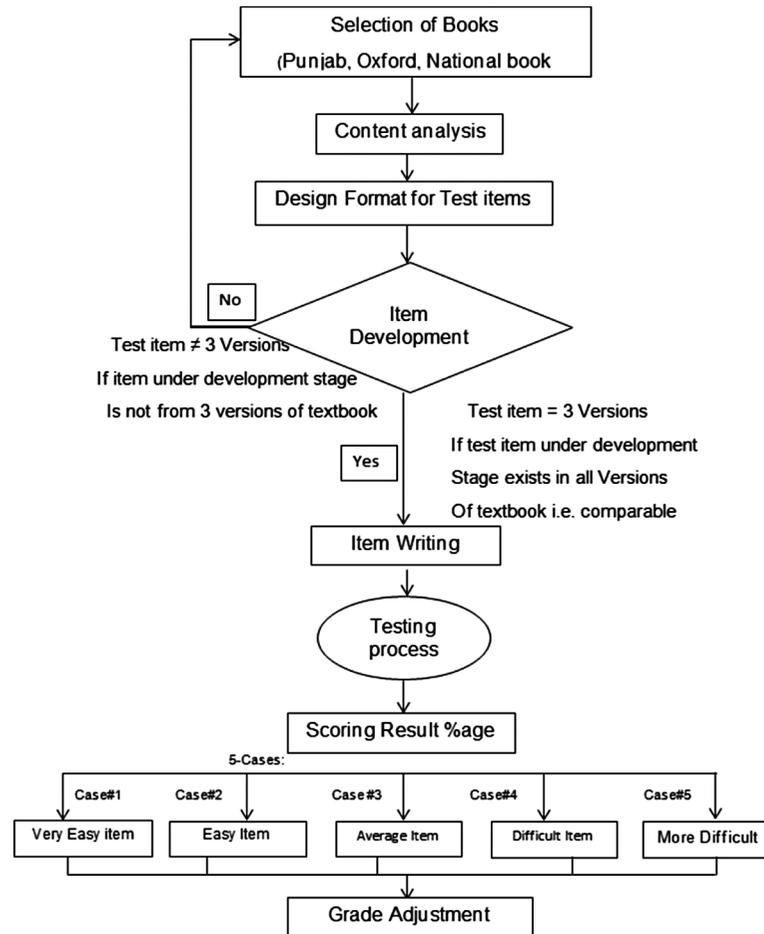
### II. Arithmetic test

- 2.1.1. Oral Mathematics
- 2.1.2. Arithmetic problems

### III. Writing test

- 3.1.1. Word spelling test.

**Figure 1. Flow diagram for development of online RAW.**



We classify test items into selected response questionnaire format and constructed response format. The former are those in which participant has to choose relevant answer from answers list i.e. True/false or multiple-choice questions (MCQ). The latter are questions in which participants present an answer after making some calculation i.e. completion items. We use selected response type questions to judge students learning in English letter reading, English word reading and spellings test, and Oral Mathematics. For assessing mathematical skills, we use constructed response format.

### 2.1. Development of achievement test

In order to develop items, we select a total of 30 textbooks from grade 1 to grade 5 curriculums such that six textbooks of English and mathematics from each grade level. We follow three version textbook styles, textbooks for all grades with detail is mentioned in Table 1.

We assign weights to test contents so that we are able to assign difficulty levels and prepare subtypes. Each grade has a total of 50 marks with reading, arithmetic, and writing tests are of 20, 20 and 10 marks each. These questions include closed item, short question answers, descriptive questions, true-false items, and multiple-choice questions. We define difficulty level for test item as mentioned in Table 2.

**Table 1. Selection of textbooks**

Classes/textbooks	National book foundation	Punjab textbook	Oxford book	Total no. of books
Class 1	2	2	2	6
Class 2	2	2	2	6
Class 3	2	2	2	6
Class 4	2	2	2	6
Class 5	2	2	2	6
Total	10	10	10	30

**Table 2. Weightage to the difficulty level**

Level	Percentage (%)
Very easy (VE)	≥90
Easy (E)	80–89
Average (A)	40–79
Difficult (D)	26–39
Very difficult (MD)	≤25

For item writing and development, we use facet design for matching test items from different books and on the basis of similar items we have selected our questions for test.

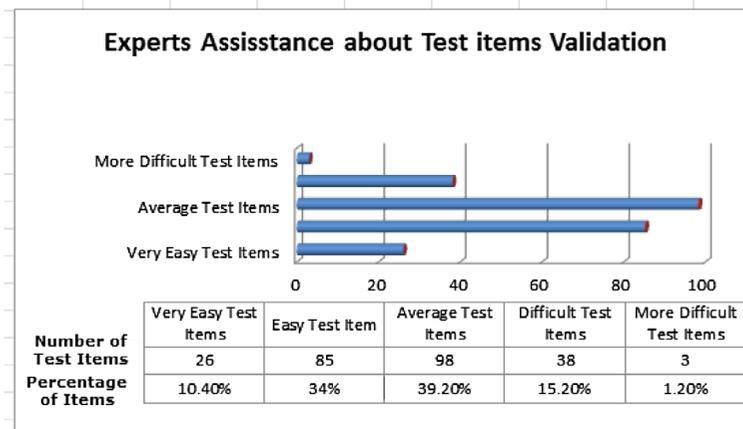
**2.2. Test items validation**

After item writing and development phase, we validate test items. First we develop and use a questionnaire to conduct analysis with the help of experts who are trained primary level teachers and psychologists. This helps us to analyze our test items through pre-testing. We then use statistical analysis tools and techniques for validation purpose.

In this section, we develop a survey from 250 test items to check item difficulty and we conclude that out of 250 test items, 26 (10.40%) test items are very easy, 85 test items (34%) are easy, 98 (39.20%) are average, 38 (15.20%) are difficult, and the rest are very difficult as shown in Figure 2.

We accept or reject our test items on the basis of results mentioned above and out of 250 test items 29 were rejected and 221 were in accepted range. We calculate reliability of 250 test items of Online RAW achievement test through SPSS statistics. Our Cronbach’s alpha (internal consistency) that we have calculated through SPSS is  $\alpha = 0.780$  as shown in Table 3.

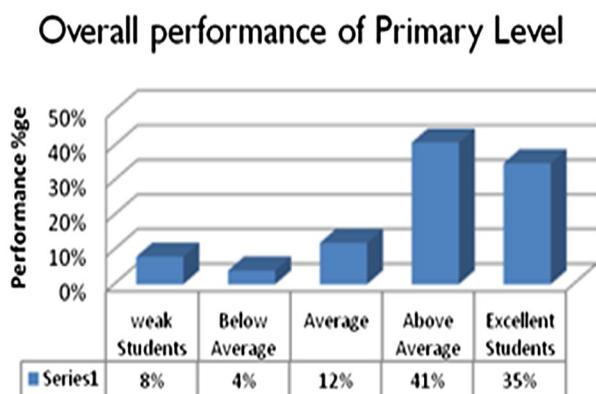
**Figure 2. Expertise judgments about test items.**



**Table 3. Reliability of online RAW**

Reliability statistics: Overall RAW test item reliability		
Cronbach's $\alpha$	Cronbach's $\alpha$ based on standardized items	No. of items
0.607	0.780	250

**Figure 3. Overall performances of primary students.**



### 2.3. Evaluation

After construction of achievement test i.e. grade-selection, content-analysis, item writing and validation, experts panel assistance through survey, we evaluate achievement test. In this section, we have compare our RAW with WRAT-4 to find which has better result and part (b) further compares both achievement tests (RAW and Wide range achievement test (WRAT)) with historical record. We select primary level students (number of students) through random sampling technique from each grade i.e. (Through grade 1 to grade 5). We check their performance according to standardized rules and criteria that we define earlier and conclude that overall performance of grade 1–5 in all tests in most of tests are above 50% and below 95%. Our students which we have tested fall in average, above average, and excellent category, detailed grade wise percentage is shown in Figure 3.

### 2.4. RAW vs. WRAT-4

We conduct a sequence of experimentations and compare our outcomes with WRAT-4 which is a well-known standardized achievement test (Wilkinson & Robertson, 2006). We select students from classes one to five and we conduct assessment using both tests i.e. RAW and WRAT-4. We compare results of RAW and WRAT to see which is performing better. We examine our result utilizing Pearson correlation of SPSS as shown in Table 4. We use Pearson correlation that ranges from -1 to +1, such that negative values indicate negative correlation and positive values indicate positive correlation as shown in Table 4.

Considering Table 4, it can be seen that zero indicates no relation existing among two groups, whereas Sig. (2-tailed) indicates *p*-value. Investigation portray that there is a critical and positive relationship exists among two accomplishment tests. Furthermore, *N* represents sample size. Value of Sign in letter reading is 0.027, 0.343 in word reading, 0.259 in spelling test, and 0.510 in oral mathematics and these all are greater than 0.05 so there exists no relationship among RAW and WRAT in above-mentioned tests but there exist a strong relationship among RAW and WRAT in math's problem test. Performance score of RAW in battery tests is 94% in letter Reading, 88% in word reading, 74% in oral mathematics, 90% in mathematical problems, and 92% in spelling tests though in WRAT students accomplish 86% in letter reading, 72% in word reading, 56% in oral mathematics, and 82% in spelling test which is not as much as RAW score, so our novel accomplishment test which is auto-mated is more effective and very much sorted out than WRAT.

**Table 4. Correlation between RAW and WRAT**

		Correlations					
		Letter reading	Word reading	Spelling	Oral maths	Maths computation	RAW vs. WRAT
Letter reading	Pearson correlation Sig. (2-tailed) N	1	0.507*	-0.174	-0.163	0.007	-0.495*
			0.023	0.464	0.492	0.977	0.027
		20	20	20	20	20	20
Word reading	Pearson correlation Sig. (2-tailed) N	0.507*	1	-0.024	0.061	0.133	-0.224
		0.023		0.922	0.798	0.575	0.343
		20	20	20	20	20	20
Spelling	Pearson correlation Sig. (2-tailed) N	-0.174	-0.124	1	0.134	0.069	-0.265
		0.464	0.922		0.574	0.771	0.259
		20	20	20	20	20	20
Oral maths	Pearson correlation Sig. (2-tailed) N	-0.163	0.061	0.134	1	0.654**	-0.157
		0.492	0.798	0.574		0.002	0.510
		20	20	20	20	20	20
Maths computation	Pearson correlation Sig. (2-tailed) N	0.007	0.133	0.69	0.654**	1	-0.637**
		0.977	0.575	0.771	0.002		0.003
		20	20	20	20	20	20
RAW vs. WRAT	Pearson correlation Sig. (2-tailed) N	-0.495*	-0.225	-0.265	-0.157	-0.637**	1
		0.027	0.343	0.259	0.510	0.003	
		20	20	20	20	20	20

\*Correlation is significant at the 0.05 level (2-tailed).

\*\*Correlation is significant at the 0.01 level (2-tailed).

**Table 5. Performance of achievement tests**

	N	Minimum	Mean	Std. error	Std. deviation
RAW vs. WRAT	20	0.00	0.5000	0.11471	0.51299
Letter reading	20	5.00	8.8500	0.30153	1.34849
Word reading	20	5.00	7.9500	0.35891	1.60509
Spelling	20	3.00	6.7500	0.38984	1.74341
Oral math	20	5.00	8.0500	0.36617	1.63755
Math computation	20	3.00	7.1500	0.55855	2.49789
Valid N (list wise)	20				

RAW has shown better result over WRAT on the grounds that analysts expected course educational programs of Pakistani reading material and build up this test for particular age bunch i.e. essential level 5–9-year students besides WRAT was developed for all age i.e. 5–94 group to check performance. Performance scores of both achievement tests are shown in Table 5.

### 2.5. Comparison of RAW and WRAT with historical data

As a second means to do evaluation of our proposal, we collect historical performance records of our subjects (students are selected for conducting WRAT-4 tests and for our proposed RAW). The results are presented in Table 6.

For conducting this evaluation, we randomly select students of grade 5 and check their performance using both achievement tests i.e. RAW achievement test and WRAT furthermore compare their performance with previous school records. We observe that the overall performance of RAW is

**Table 6. Comparison results of RAW vs. WRAT**

Tests and subtests	RAW scores (%)	WRAT scores (%)
Letter reading test	94	86
Word reading test	88	72
Oral math's	74	56
Arithmetic problems	90	72
Spelling test	92	82

**Table 7. Achievement tests vs. previous students record**

	1	2	3	4	5	6	7	8	9	10
RAW (%)	92	82	88	82	94	70	82	84	86	84
WRAT (%)	76	80	58	78	60	64	74	66	72	82
Historic results (%)	85	79	67	78	94	71	80	80	71	83

84%, Historical record is 80.5%, and WRAT is 71%. There is a 3.5% difference among RAW and historical results but 9.5% difference among WRAT-4 and historical record. We present comparison of RAW vs. WRAT in Table 7 which is a 11×3 contingency matrix.

### 3. Results and discussion

We find a number of standardized achievement tests but majority of them are not computerized and virtually all of them are region specific. Our work is significant in the sense that the test is automated and our main focus is to design this achievement for specific age group.

In test item validation, we conducted a survey from 250 test items and accept/reject items on the basis of item difficulty level and come to conclusion that out of 250 test items 221 are in acceptable region and 29 are in rejected region. Reliability coefficient is basically the co-relation among two sets of test marks (Cortina, 1993). Reliability can be checked through different ways i.e. Cronbach alpha and Kuder–Richardson-20 formula of reliability and the reliability coefficient that could be computed through this way is called internal consistency of reliability. We test reliability through SPSS software i.e.  $\alpha = 0.780$  as shown in Table 6. First we checked the performance of students through a computerized way i.e. RAW and then also checked the performance of same group of students through paper-pencil assessment way i.e. WRAT. We applied Pearson correlation of SPSS on all the results achieved through both achievement tests and found that there exists a significant difference among RAW and WRAT. We have concluded through evaluation part I i.e. RAW vs. WRAT and found that the performance of letter reading is 94%, 88% in word reading, 74% in oral mathematics, 90% in mathematical problems, and 92% in Spelling test achieved through RAW achievement test, whereas in WRAT test the performance of students in letter reading is 86%, 72% in word reading, 56% in oral mathematics, 72% in mathematical problems, and 82% in spelling test. On the basis of these results, we conclude that RAW score is providing a better reflection of the students' achievements.

### 4. Related work

Perrone (2013) explains “classroom level achievement tests: An essential part of the second language learning and teaching processes” and demonstrates that an achievement test is mainly used in building classroom-level assessments and is particularly intended with reference to course goal and learning objectives that are set according to curriculum, and researchers express that achievement test measures student performance in a particular domain and for particular grade level. The researchers review three main resources that can be followed as a course outline i.e. books, syllabus, and course objectives. Investigator also describes the limitation and benefits of the achievement test as indicated in Perrone (2013). Another similar study (Zunaira Fatima, 2015) has exposed the development and Rasch analysis of an Achievement test for master level in the subject of philosophy

of education. Scholars develop 60 test items and perform item analysis through Rasch analysis technique and refine properties of items using latent continuum and concluded probability of test items utilizing the Rasch model. In 2015, Wiley explains “Comparison of the Gates Reading Survey and the Reading Section of the WRAT” and the major purpose of WRAT is to measure personality in three major areas i.e. spelling, reading, and arithmetic problems to evaluate individual’s strengths and weaknesses. The main purpose of the WRAT is to find the relationship between a group reading test and an individual test (Fortenberry & Broome, 2015).

Researchers adopted a systematized and structured approach to evaluate and present an overview of educational, psychological, and academic profile of the individual (Sax, 1997). Achievement tests i.e. Alabama reading and mathematics test (ARMT) develop in Alabama state to check performance in English word reading and mathematical problems for grade 3–8 level students. Researchers utilizing paper-pencil assessment approach to judge performance level (Kenneth & Meier, 2013). Ohio Achievement test (OAT) is an institutionalized way created in Ohio state and that is intended to judge strengths and weaknesses of understudies and judge English proficiency of English language acquisition for 6–8 grade students and subcategory of Ohio is Ohio graduate testing (OGT) and was intended to check performance in science and social studies of tenth grade students. In 2009, work on writing test was canceled and social science was introduced. Researchers check the performance of students and reading capability utilizing paper appraisal method (Ohio, 2009).

There are numbers of achievement test developed in different states i.e. ARMT in Alabama, SBA in Alaska, AIMS in Arizona, STAR in California, CSAP in Colorado, FCAT in Florida, ISAT in Idaho, MSA in Maryland, MAP in Missouri, MONTCAS in Montana, NMSBA in Mexico, WRAT in Wilmington, NYSTP in New York, OAT in Ohio, and WCAP in Washington, but there is no development of standardized achievement test with battery and application. We develop an achievement test and propose technique for test items development to check performance of students.

## 5. Conclusion

We develop an achievement test and name it “Online RAW test” and we create a battery for current test. We also develop an application for this achievement test so that educational psychologist and therapist can easily calculate performance score of students efficiently through a computerized way to judge students learning disorder. We develop test items from primary learner textbooks to check student performance and in addition for course improvement through expert’s opinion.

As an outlook, we plan to extend this achievement test for 6 to 10 grade students. We also plan to utilize psychometrical principles and we also intend to add multiple subtests in battery.

### Funding

The authors received no direct funding for this research.

### Author details

Maria Hanif<sup>3</sup>

E-mail: [mariyahanif90@gmail.com](mailto:mariyahanif90@gmail.com)

Tamim Ahmed Khan<sup>1</sup>

E-mail: [tamim@bui.edu.pk](mailto:tamim@bui.edu.pk)

ORCID ID: <http://orcid.org/0000-0002-8209-6100>

Uzma Masroor<sup>2</sup>

E-mail: [uzmamasroor@yahoo.com](mailto:uzmamasroor@yahoo.com)

Amira Amjad<sup>2</sup>

E-mail: [psychologist.ccp@gmail.com](mailto:psychologist.ccp@gmail.com)

<sup>1</sup> Department of Software Engineering, Bahria University, Islamabad, Pakistan.

<sup>2</sup> Department of Professional Psychology, Bahria University, Islamabad, Pakistan.

### Citation information

Cite this article as: Development of online RAW achievement battery test for primary level, Maria Hanif, Tamim Ahmed Khan, Uzma Masroor & Amira Amjad, *Cogent Education* (2017), 4: 1290332.

### References

- American Educational Research Association. (1999). *Standards for educational and psychological testing*. Washington, DC. Retrieved from <http://www.jstor.org/stable/1163515>
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78, 98–104. <http://dx.doi.org/10.1037/0021-9010.78.1.98>
- Fortenberry, W. D., & Broome, B. J. (2015). *Comparison of the gates reading survey and the reading section of the wide range achievement test* (pp. 66–68).
- Gay, L. (1996). *Educational research*. Upper Saddle River, NJ: Prentice Hall.

- Kenneth, J., & Meier, A. R. (2013). Alabama achievement test. In J. Kenneth & A. R. Meier (Eds.), *Politics of African-American education: Representation, partisanship, and educational equity* (illustrated ed., p. 234). Cambridge: Cambridge University Press.
- Ohio. (2009). *ODE-Resources for Ohio Achievement Tests Ohio Department Idaho State Idaho State Department of Education-Assessment Idaho standard of achievement test (I-SAT) Education*, Idaho State department Idaho State Department Idaho State.
- Perrone, M. (2013). Classroom achievement tests: An essential part of the second language learning and teaching processes. *TESOL & Applied Linguistics*, 3(1).
- Sax, G. (1997). *Principles of educational and psychological measurement and education*. Washington, DC: International Thomson.
- Wilkinson, G. S., & Robertson, G. J. (2006). *Wide range achievement test*. Lutz, FL: Psychological Assessment Resources.
- Zunaira Fatima, S. H. (2015). Development and Rasch analysis of an achievement test at master level (philosophy of education). *Pakistan Journal of Commerce and Social Sciences*, 9, 269–281.



© 2017 The Author(s). This open access article is distributed under a Creative Commons Attribution (CC-BY) 4.0 license.

You are free to:

Share — copy and redistribute the material in any medium or format  
Adapt — remix, transform, and build upon the material for any purpose, even commercially.  
The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:

Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made.  
You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.  
No additional restrictions

You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.



**Cogent Education (ISSN: 2331-186X) is published by Cogent OA, part of Taylor & Francis Group.**

**Publishing with Cogent OA ensures:**

- Immediate, universal access to your article on publication
- High visibility and discoverability via the Cogent OA website as well as Taylor & Francis Online
- Download and citation statistics for your article
- Rapid online publication
- Input from, and dialog with, expert editors and editorial boards
- Retention of full copyright of your article
- Guaranteed legacy preservation of your article
- Discounts and waivers for authors in developing regions

**Submit your manuscript to a Cogent OA journal at [www.CogentOA.com](http://www.CogentOA.com)**

