

# Self-Assessment in Academic Writing: A UAE Case Study

---

David Litz

UAE University

## Abstract

The main purpose of this study was to examine the relationship between student ratings and teacher ratings on a university-level writing assignment in order to investigate the extent to which students have the ability and wherewithal to accurately and reliably assess themselves on a finished writing product. The Pearson Product Moment coefficient was used to determine whether or not there was a statistical correlation between student scores and teacher scores while the Intra-Class Coefficient and Spearman Brown Prophecy formula were used to determine the degree of agreement between raters as well as amongst all of the raters for an average reliability score. In this case the results were very promising as it was found that student and teacher scores correlated very highly and demonstrated a strong degree of agreement. This suggests that self assessment may be used to assist students in this particular context to better understand the conventions of English writing and ultimately improve their overall writing ability.

## Introduction

### ***Learner-centered assessment in English as a second/foreign language***

During the past decade, movements towards learner-centered language instruction as well as a renewed interest in the authenticity, interactive pedagogy, and the overall impact of language assessments have gradually come together to generate an increased awareness in expanding the role of the learner and test-taker in the teaching and learning process (Bachman, 2000). Learner-centered teaching has focused on the types of activities and tasks with which learners interact in the language classroom, the promotion of self-access or independent learning centers, and the use of a wide variety of multi-media computer-based learning programs (Bachman, 2000). Moreover, teachers have been encouraged to take a holistic view towards their students' education whereby students are taught to take responsibility for their own learning and think independently both inside and outside the classroom (Ekbatani, 2000; Stiggins, 1997). Unfortunately, appropriate assessment procedures in which learners are not only test takers but also active participants in the assessment process have often been absent from learner-centered teaching in the past. This has consequently led to renewed criticism of standardized tests and traditional testing techniques. It has also led to an interest in making language assessments more authentic and interactive, and for facilitating a positive or favorable impact of tests or assessments on test-takers (Ekbatani, 2000; Stiggins, 1997).

## Self assessment: what the literature says

### ***Self assessment: what is it?***

The educational system in the United Arab Emirates (UAE) has recently come under a great deal of criticism by the country's rulers. It is expected that a number of changes will be implemented in the K-12

and post-secondary systems in the coming years as part of the UAE development and educational reform plan that was recently announced. One reason for the failure of the educational system could be the reliance on traditional testing and assessment methods.

Assessment in many Middle Eastern classrooms has traditionally focused on teacher-designed “paper and pencil” tests interjected at the end of a unit of study to determine the extent of learning and to ultimately assign a grade. This type of assessment generally focuses on memorization and basic skill development and has led to mounting disapproval and condemnation with standardized tests and traditional testing procedures. It has also led to a call for the increased use of assessment approaches that go beyond the traditional format to include strategies like alternative or authentic assessment (Barlow & Coombe, 2001; Stiggins, 1997). These alternative or authentic assessment approaches reflect a shift in emphasis from assessment as a selection procedure to one of illumination, accountability and genuine learning.

One of the most interesting and effective forms of authentic assessment and testing that heightens learner awareness of personal strengths and weaknesses and promotes genuine language acquisition is self-assessment. Nunan (1988) was one of the first proponents of self-assessment. He argued that this type of assessment can be an effective supplement to teacher assessment and that self-assessment can be an integral means of developing critical self-awareness of what it is to be a learner as well as the skills of learning how to learn. Dickinson (1987) also suggested that self-assessment is an important part of pedagogy because a) assessment leading toward evaluation is an important objective in its own right and training learners in this is beneficial to their overall learning b) assessment is a necessary part of self-direction and c) in instructional programs involving many students, the assessment demands made by the students are significant and self-assessment is one way of alleviating teachers’ heavy assessment burdens.

One of the primary concerns with self-assessment has centered on the reliability of self-assessment techniques used as a basis for making decisions such as selection, grading, ranking or certification as well as learners’ objectivity and capacity to assess themselves. Nunan (2007: 148) states that this “criticism misses the point to some extent, which is to involve learners in their own learning processes”. This argument is further supported by Cram (1995: 282, cited in Nunan (2007)), who suggests that “[t]he major purpose of self-assessment is to provide the opportunity for learners to develop an understanding of their own level of skill, knowledge or personal readiness for a task in relation to their goals”.

## **Self Assessment in Writing**

A number of theorists (e.g. Farr and Tone, 1994; O'Malley and Pierce, 1996; Thomas & Barksdale-Ladd, 2000) have all argued that self-assessment can be an effective means of improving students’ writing. These authors contend that self-assessment promotes self-reflection, self-esteem and self-confidence which ultimately improve students’ writing (El-Koumy, 2004). Other practitioners (e.g. Andrade & Du, 2005; Cunningham and Cunningham, 1987; Kirszner and Mandell, 1989; Lebauer and Scarcella, 1992; Radecki and Swales, 1988; Ross et al., 1999) have expanded upon this research and have demonstrated that students can actually play a key role in assessing the final product of their writing. They

claim that this type of self-assessment helps students to master English writing conventions which can, in turn, enhance their writing. They also claim that self-assessing the outcomes of writing helps students to evaluate the degree of attainment of the writing-program goals. (El-Koumy, 2004: 1).

The aim of this particular study was to describe a situation in which self-assessment was introduced to ESL students in a writing course in the University General Requirements Unit (UGRU) of UAE University

(UAEU) as well as to present the findings of several statistical analyses that were conducted in order to investigate the degree of correlation and inter-rater reliability between teacher-rating and student self-assessment on a finished writing product. While self-assessment is not currently used for university-level writing projects at UGRU, a high degree of correlation and inter-rater reliability between the teachers and students in this particular study would demonstrate that each group is rating the completed project in a similar fashion. It would also lend support to the view that students have the ability to accurately assess a final product of writing and suggests that students may, in fact, have gained a richer understanding of their writing ability in addition to a deeper appreciation for English writing conventions and the desired outcomes and expectations of this particular writing task. An insignificant correlation and low degree of inter-rater reliability, on the other hand, would suggest that self-assessment did not work in this instance and might not be appropriate in this educational context.

## **Methodology**

### ***Purpose and Rationale***

The primary goal of this particular study was to assess reliability in student self-assessment of the final product of their writing. In order to conduct such a study, it was decided to conduct some simple correlation coefficient analyses between student self-assessment scores and teacher-rating scores on an actual graded university-level writing assignment.

It is likely that a correlational study of this nature will support or dispute the view that students have the ability and wherewithal to accurately and reliably assess themselves on a finished writing product and that this type of assessment should, therefore, be integrated into the writing curriculum and utilized for more writing tasks. If, for example, the correlation between teacher-rated scores and student self-assessment scores is found to be significant then writing teachers at UGRU should be allowed to incorporate self-assessment into their classrooms and assessment regimens as this would be an effective evaluation innovation that is consistent with many of the student-centered core values and mission of UGRU. Conversely, if teacher-rated scores and student self-assessment scores are found to correlate slightly or the results of the experiment are either statistically insignificant or meaningless then it would seem as though this type of authentic assessment may not be appropriate for UGRU writing courses at the present time.

### ***Learning Context***

As mentioned previously, this particular study was conducted at UAE University within the English Program of UGRU. The English Program's goal is to ensure that all students entering the university faculties are equipped with the necessary study skills and language proficiency to be successful university students. Students who do not possess an IELTS score of 5 are placed into one of three levels depending on their CEPA score:

- Level 1 is an elementary course that: reviews basic grammatical structures; expands familiarity with common vocabulary; introduces basic reading and listening skills; introduces basic paragraph writing skills; encourages speaking in the classroom context; focuses on how to learn; and encourages independent study.
- Level 2 is a pre-intermediate course that: consolidates and expands upon basic grammatical knowledge; focuses on high frequency vocabulary; develops reading and listening skills, practices taking notes from short lectures; develops paragraph writing and introduces short

essay skills; encourages speaking in the classroom context develops study skills; encourages critical thinking skills; and stresses the importance of independence study.

- Level 3 is an intermediate level course that: prepares students for survival in an academic context by expanding upon grammatical knowledge of common structures; focusing on general academic vocabulary; developing reading skills with reference to challenging texts; practices note-taking from lectures; develops academic essay writing skills; encourages speaking in an academic context; reinforces the importance of study skills, critical thinking skills and independent learning in an academic context.

All of the courses last for one semester and students will typically study integrated skills with one teacher for 8 hours a week and writing skills with another teacher for an additional 4 hours a week. The integrated skills component makes up 70% of a student's total English score while the writing component makes up 30%.

### ***Subjects***

The subjects for this study consisted of 2 writing teachers and 35 Level 3 female writing students. All of the student subjects were volunteers in two classes. These classes were both 600-level (repeater) classes. All of the students were familiar with scoring rubrics, but had never assessed themselves on a finished piece of university-level work. While these students had completed a similar type of writing project the previous semester, they had not assessed themselves so it was felt that their prior exposure would not affect the results. It was agreed that it would be beneficial to use students of comparable ability in order to decrease the likelihood of varying proficiency levels having an undue influence on the outcome of the study. Prior to undertaking the study, all the volunteers were also asked to read and sign a 'permission statement' in Arabic and English. This permission statement was based on The University of Calgary's Human Subjects Consent Form.

### ***Materials and Design***

For the purpose of this study, a number of formal instruments were utilized in order to acquire sufficient data on the variables discussed above. These included the actual writing task and the scoring rubric that was used by the students and teachers for assessing the final writing product. A complete timeline and sequence of events that was utilized to introduce the students to the task as well as the scoring rubric, write the first and second draft of the essay and score the assignments is outlined in Figure 1:

**Figure 1: Sequence of stages by class period.**

<b>Class Period 1</b>	Introduce assignment and scoring rubric. Hand out, read and discuss qualities of model essay. Hand out and examine scoring rubric – discuss different criteria and gradations of quality.
<b>Class Period 2</b>	Prewriting, (e.g. outline / plan, brainstorming etc.)
<b>Class Period 3</b>	Students write first draft of essay.
<b>Class Period(s) 4 +</b>	Teacher gives each student feedback for the first draft. Students write final drafts. Students self-assess final drafts with scoring rubric.

	Teacher(s) assess final drafts with scoring rubric (outside of class time).
--	---

### **The Writing Task**

The students' writing project was a 5-paragraph agree/disagree opinion essay. The students were expected to have an introduction paragraph, 3 body paragraphs and a conclusion paragraph. They were also expected to use quotations and/or references and write a minimum of 250 words.

### **The Scoring Rubric**

The scoring rubric that was used for this study is the same one used for all criterion-referenced scoring of Level 3 writing projects at UGRU. It was developed by the UGRU Writing Committee and has been in use in its current form for several years. It had been trialed and vetted on numerous occasions in the past. It consists of a fixed measurement scale and sets of criteria that describe the characteristics or gradations of quality for each score point or range. Essentially, students can score a maximum of 5 points for each of: task fulfillment; vocabulary usage; organization, cohesion and style; grammar and form. The total project score is awarded out of 20 possible points (See Appendix 1). No mark is awarded for the writing process, but the students generally undertake a plan, and several drafts. General comments that pertain to the writing process are also provided by the teacher.

The students had all been provided with this rubric by their teachers prior to working on their writing project. In this way they were all familiar with the descriptive criteria and had an opportunity to gain an appropriate understanding of the task requirements as well as their teachers' expectations. A special scoring training session was also provided to the students participating in the study. Both of the teachers that were used for this study were experienced Level 3 writing teachers with training in the use of scoring rubrics. They were familiar with the scoring rubric and had participated in a rigorous calibration/banding process that preceded the study. In addition, inter-rater reliability was ensured through a blind scoring session prior to the study. Each teacher's scores were totaled and analyzed statistically with the Intra-class Correlation Coefficient (ICC). The ICC is designed to provide the degree of agreement among different raters. The range of the ICC may be between 0.0 and 1.0 and the ICC will be high when there is little variation between the scores given to each item by the raters (e.g. if all raters give the same, or similar scores to each of the items). In this case the ICC inter-rater reliability score for the banding/calibration scores was estimated to be 0.88, which is quite high.

### **Data Collection**

A protocol for the administration of all instruments was designed to insure that procedures would be standard throughout data collection. Data for the study was collected after the students had completed their writing projects. Essentially, the results of the writing task were scored by the students as well as the teachers and both sets of raw scores were statistically analyzed and compared in order to check for a meaningful, significant and reliable correlation between self-assessment and teacher-rating for this particular writing task. All of the subjects were told that they were participating in a study and that the self-assessed scores for the project would have no effect on their overall course grade, but they were informed about the nature and purpose of the experiment. The final student ratings as well as teacher ratings for the completed projects can be seen in Appendix 2.

## Results and Discussion

### *Descriptive Statistical Analysis*

The normal distribution of data is a crucial factor for many statistical measures and subsequently is one of three basic assumptions that underlie Pearson's Product Moment Correlation Coefficient - along with *independence* and *linearity* (Brown, 1996). In order to examine the normality of the distribution of scores, the means (averages), medians and modes as well as standard deviations and variance were calculated. Values are shown in Table 1:

**Table 1: Descriptive statistics.**

Statistical Measurements	Average	Median	Mode	Standard Deviation	Variance
Student Scores	14.85714	15	16	2.031268	4.12605
Teacher A Scores	14.35714	14	12	1.853341	3.434874
Teacher B Scores	14.84286	15	16	1.901127	3.614286

In this case the results of the descriptive statistics indicate that each set of data has fairly similar and acceptable values and this is representative of a normal distribution of scores or central tendencies. It is also a superficial indicator that these measures (teacher scores and student scores) may be closely related to each other in terms of the ability they assess. *Independence*, on the other hand, requires that each score pair is distinct from all other score pairs, and care was taken in inputting and charting the data to prevent any systematic association between pairs of scores during the statistical analysis. Finally, several *scatterplots*, which provide a representation of sets of scores for visual comparison, were produced to determine if the relationship between all the sets of scores was in fact linear, and the resultant graphs demonstrate suitable linearity (See Appendix 3).

### *Inferential Statistical Analysis*

While the distribution of scores, particularly the means and standard deviations, were within very acceptable ranges and comparable to one another, more thorough analyses are required in order to determine if there is actually a true statistical relationship or correlation between the student scores and teacher ratings. The first statistical measure that was calculated was the *Pearson's Product Moment* correlation coefficient (Pearson's *r*). We use this particular formula to gauge the degree of correlation between two or more variables. Coefficient values may range from +1.0 (a perfect positive correlation) to -1.0 (a perfect negative correlation) (Brown, 1996). In this study, the variables refer to the student ratings, teacher-A ratings and teacher-B ratings and the resultant coefficients provide information regarding the extent to which the sets of scores vary in relation to each other. Setting the level of significance at .001( $p < 0.001$ ), the student scores and teacher-A scores had a correlation index of .963 while the student scores and teacher-B scores had a correlation index of .961 (teacher-A and teacher-B scores had a correlation index of .949) (see Table 2). This would seem to indicate, therefore, that there is indeed an extremely high or almost exact correlation between the student self-assessed scores and those given by the teachers.

**Table 2: Inferential statistics.**

Statistical Measurements	Pearson's Product Moment	Intra-class correlation coefficient (ICC)	Spearman-Brown Prophecy
Student/ Teacher-A Scores	.963	.9363	.9573
Student/ Teacher-B Scores	.961	.9603	.9573
Teacher-A/ Teacher-B Scores	.949	.9182	.9573

The second component to the inferential statistical analysis consisted of the calculation of the *Intra-class correlation coefficient (ICC)* and the *Spearman Brown Prophecy* because there is always a possibility that the students' and teachers' scores could be highly correlated with one another, but show little actual agreement. These measurements can be used to gain a clearer picture of the extent to which the students' and teacher' scores actually agreed as they provide us with the degree of agreement, homogeneity or consensus between raters (ICC) as well as amongst all of the raters for an average reliability score (*Spearman Brown*). In this case the students and teacher-A had an intra-class coefficient score of .9363 while the students and teacher-B had an intra-class coefficient score of .9603. These scores indicate that there was a very strong intra-class coefficient correlation or degree of agreement between the students' scores and the teachers' scores. Interestingly enough, teacher-A and teacher-B had an intra-class coefficient of .9182. While still very high, this was the lowest intra-class correlation coefficient of all the sets of scores. The final correlational analysis that was run was the *Spearman Brown Prophecy* coefficient in order to gain a measurement of the reliability of all the raters averaged together. In this case the correlation was .9573 which is very strong at the .001 level of significance.

### **Discussion**

The outcomes of this study were promising as they indicated that there was a significant and meaningful correlation between student scores and teacher scores in addition to a strong degree of agreement or reliability between the individual raters' scores for a finished piece of writing. This would seem to suggest, therefore, that self-assessment may actually be assisting the students to improve and enhance their understanding of task expectations, their own overall writing ability and the conventions of English writing as well as the skill of utilizing holistic scoring rubrics for the purpose of rating themselves. Consequently, teachers might want to consider the use of self-assessment to a larger extent in writing classrooms at UGRU in the future because student self-rating is not only a useful strategy which helps the individual learner move toward autonomy in self-monitoring and self-directed language learning, but it also gives us a clearer picture and indication of our students' progress, especially when they are directly engaged in their own learning and assessment in a formal academic context. As Andrade et al. (2008: 9) suggest: "[b]y involving students in the assessment process in these ways, teachers can blur the distinction between instruction and assessment and transform classroom assessment into a moment of learning".

### **Limitations and recommendations for future research**

The results of this study were limited by the sample size, characteristics and homogeneity of the subjects and length of the study and the distributions of scores may have been influenced by these factors. All of the data, for example, was collected over the course of one semester from female subjects that were between the ages of 18-21. These subjects also came from two Level 3 repeater (600) sections. Further longitudinal and large-scale studies are needed with students of different sexes, ages, levels and ranges of ability. Moreover, this particular study presented findings from a particular

homogeneous Emirati (UAE) context. Additional research needs to address similar issues in different EFL/ESL contexts and the degree to which culture has an impact on self assessment choices, preferences and practice. These studies might examine the relationship between self-assessment and specific learner needs, behaviours, strategies, expectations or even students' particular cultures of learning.

Another limitation of this study was the fact that it solely focused on the relationship between the final student ratings and teacher ratings for a particular writing assignment. Recent research by Andrade and Boulay (2003); Andrade and Du, (2005); and Andrade et al. (2008) has demonstrated that students who were actively involved in the actual scoring rubric creation process in addition to self-assessing themselves tended to improve their average writing scores and gain better overall scores than students who did not assess themselves or merely used a stand-alone rubric for self-assessment. More research is certainly needed to further investigate whether students' overall writing ability improves over time as a result of being involved in self-assessment classroom activities as well as to study the degree to which student involvement in the scoring rubric construction process plays a significant role in improving writing ability.

A third limitation of the study is the fact that it concentrated on the final scores of a finished writing product. There is no question that comparing the teacher and student scores for each of the items or categories of the rubric may have provided even more detailed information about self-assessment and studies of this nature will need to be conducted in the future. Moreover, a number of researchers (e.g., Angelo and Cross, 1993; Dickinson, 1987; Johnston, 1987; Keh, 1990; Lehr, 1995; Reid, 1988) have argued that studies that focus on a final writing product often ignore the cognitive processes that students go through in producing their writing and have called for self-assessing the processes of writing (El-Koumy, 2004). Advocates of this view claim that this type of self-assessment is better at helping students to bring the writing processes

into conscious awareness which can, in turn, improve their writing outcomes. They also claim that self-assessing the processes of writing helps students to tailor strategies for the purpose of writing, and to select appropriate writing strategies that suit the writing task. (El-Koumy, 2004: 1).

However, this type of self-assessment has also been criticized for neglecting the conventions and standards of typical English writing (El-Koumy, 2004; Myers, 1997). In addition, it has been suggested that it may actually interfere with the normal processes of writing and – even if it occurs at the end of writing – there is the possibility that students might forget or possibly incorrectly recall the strategies they used before and during a particular writing task (El-Koumy, 2004; Hayes and Flower, 1983). It would seem, therefore, that this is a hotly contested issue and more studies are definitely going to be required that compare the effect of self-assessment of writing processes versus the self-assessment of actual final products of EFL students' writing in the future.

A final limitation of the study is related to the use of more sophisticated statistical operations such as the ICC and *Spearman Brown Prophecy*. While these formulas do demonstrate a degree of inter-rater reliability and average agreement for all raters, there are several drawbacks that researchers need to consider. First, ICCs are highly dependent on the trait variance within the population for which it is measured. This can complicate comparisons of ICCs measured in different populations, or in generalizing results from a single population and the same statistical instrument may also be judged "reliable" or "unreliable," depending on the population in which it is assessed (Uebersax, 2007). Second, ICCs combine two ways in which raters differ: (1) *association*, which concerns whether the raters understand the meaning of the trait in the same way, and (2) *bias*, which concerns whether some raters' mean ratings are higher or lower than others. If a goal is to give feedback to raters to improve future ratings, one should distinguish between these two sources of disagreement (Uebersax, 2007). Lastly, ICCs



discount the fact that there is still a basic unit to evaluate rating consistency: the number or percent of agreements on each rating category. Raw agreement is simple, intuitive, and significant and there is really nothing wrong with reporting reliability with an ICC in conjunction with raw agreement rates (Uebersax, 2007).

## Conclusion

This was a relatively small-scale study and additional studies of this nature will undoubtedly be required in the future. Nevertheless, the results were extremely favourable and certainly support the view that students are capable of accurately assessing themselves on a finished writing task as the students' ratings and the teachers' ratings displayed a high degree of correlation and inter-rater reliability. This would seem to suggest, therefore, the students and teachers were rating in a similar manner and that self-assessment may be used in this pedagogic context (and possibly other higher educational contexts in the UAE) as this type of novel assessment practice is not only more genuine, student-centered, and meaningful than the traditional types of assessments that are typically utilized in many post-secondary institutions throughout the Middle East, but it is also helpful in assisting students to master English writing conventions, gain a better understanding of scoring rubrics as well as the desired writing outcomes, and ultimately improve their overall writing ability.

## References

- Andrade, H. Boulay, (2003). The role of rubric-referenced self-assessment in learning to write. *Journal of Educational Research*, 97(1), 21-34.
- Andrade, H., & Du, Y. (2005). Student perspectives on rubric-referenced assessment. *Practical Assessment, Research and Evaluation*, 10(3), 1-11.
- Andrade, H., Du, Y. & Wang, X. (2008). Putting rubrics to the test: the effect of a model, criteria generation, and rubric-referenced self-assessment on elementary school students' writing. *Educational Measurement: Issues and Practice*, 27(2), 3-13.
- Angelo, T. & Cross, K. (1993). *Classroom assessment techniques: A handbook for college teachers*. San Francisco: Jossey-Bass.
- Bachman, L. (2000). Foreword. In G. Ekbani & H. Pierson (eds.), *Learner-directed assessment in ESL* (pp. ix-xii). Mahwah, NJ: Lawrence Erlbaum Associates.
- Barlow, L. & Coombe, C. (2001). Implementing alternative assessment in the Arabian Gulf. In Z. Syed and D. Heuring (eds.). *Tools of the trade: Teaching in the Gulf: Proceedings of the First Annual Teacher-to-Teacher Conference, Abu Dhabi, 3-4 May 2000*. Abu Dhabi: Military Language Institute. (pp. 189-199).
- Brown, J. (1996). *Testing in language programs*. Upper Saddle River, NJ: Prentice Hall Regents.
- Cram, B. (1995). Self-assessment: from theory to practice. Developing a workshop guide for teachers. In G. Brindley (ed.) *Language assessment in action* (pp. 271-306). Sydney: Macquarie University, National Centre for English Language Teaching and Research.
- Cunningham, P. & Cunningham, J. (1987). Content area reading-writing lessons. *The Reading Teacher*, 40, 506-512.
- Dickinson, L. (1987). *Self-instruction in language learning*. London: Cambridge University Press.

- Ekbatani, G. (2000). Moving toward learner-directed assessment. In G Ekbatani & H. Pierson (eds.), *Learner-directed assessment in ESL* (pp. 1-11). Mahwah, NJ: Lawrence Erlbaum Associates.
- El-Koumy, A. S. A. (2004). *Effect of self-assessment of writing processes versus products on EFL students' writing*. Paper presented at the Tenth EFL Skills Conference at the American University of Cairo, Center for Adult and Continuing Education, Cairo, Egypt. (ERIC Document Reproduction Service No. ED 490 559). Retrieved June 18, 2008, from ERIC.
- Farr, R. & Tone, B. (1994). *Portfolio and performance assessment: Helping students evaluate their progress as readers and writers*. Orlando: Harcourt Brace and Company.
- Hayes, J. & Flower, L. (1983). Uncovering cognitive processes in writing. In P. Mosenthal, L. Tamor & S. Walmsley (eds.), *Research on writing: Principles and methods* (pp. 207-220). New York: Longman.
- Johnston, P. (1987). Assessing the process, and the process of assessment in the language arts. In James R. Squire (ed.), *The dynamics of language learning: Research in reading and English* (pp. 335-357). Urbana, Illinois: ERIC Clearinghouse on Reading and Communication Skills.
- Keh, C. (1990). Feedback in the writing process: A model and methods for implementation. *ELT Journal*, 44(4), 294-304.
- Kirsznner, L. & Mandell, S. (1989). *The Holt handbook*. Fort Worth: Holt Rinehart and Winston.
- Lebauer, R. & Scarcella, R. (1992). *React: Multicultural reading-based writing modules*. Englewood Cliffs, NJ: Prentice Hall.
- Lehr, F. (1995). *Revision in the writing process*. Bloomington, IN: ERIC Clearinghouse on Reading English and Communication.
- Myers, S. (1997). Teaching writing as a process and teaching sentence-level syntax: Reformulation as ESL composition feedback. *TESL-EJ*, 2(4), 1-16. Retrieved June 15, 2008 from: <http://tesl-ej.org/ej08/a2.html>
- Nunan, D. (1988). *The learner-centered curriculum*. London: Cambridge University Press.
- Nunan, D. (2007). *Task-based language teaching*. Cambridge: Cambridge University Press.
- O'Malley, J. & Pierce, L. (1996). *Authentic assessment for English language learners: Practical approaches for teachers*. Reading, MA: Addison-Wesley.
- Radecki, P. & Swales, J. (1988). ESL student reaction to written comments on their written work. *System*, 16, 355-365.
- Reid, J. (1988). *The process of composition*. (2<sup>nd</sup> ed.). Englewood Cliffs, NJ: Prentice Hall.
- Ross, J. A., Hogaboam-Gray, A., & Rolheiser, C. (2002). Student self-evaluation in grade 5-6 mathematics effects on problem solving achievement. *Educational Assessment*, 8(1), 43-59.
- Stiggins, R. (1997). *Student-Centered Classroom Assessment*. Camden, NJ: Merrill.
- Thomas, K. & Barksdale-Ladd, M. (2000). Metacognitive processes: Teaching strategies in literacy education. *Reading Psychology*, 21(1), 67-84.
- Uebersax, J. (2007). *Intraclass correlation and related methods*. Retrieved June 15, 2008 from: <http://ourworld.compuserve.com/homepages/jsuebersax/icc.htm>

## Author

**David Litz** is currently a lecturer at UAE University and he has previously taught in South Korea and Canada. He is presently working on a doctorate in education from the University of Calgary and his professional interests include testing and assessment, educational administration and teacher training.

## Appendix 1

### Project Scoring Rubric

	5/5	4/5	3/5	2/5	1/5
Task Fulfillment	<ul style="list-style-type: none"> <li>*Excellent understanding of the topic</li> <li>*Excellent/original reasons</li> <li>*Well chosen quotes, examples, and results for support</li> <li>*Clear thesis</li> <li>*Ideas are original and fluent</li> </ul>	<ul style="list-style-type: none"> <li>*A good understanding of the topic</li> <li>*Good/clear reasons</li> <li>*Enough quotes, examples, and results</li> <li>*Thesis is present</li> </ul>	<ul style="list-style-type: none"> <li>*A basic answer to the question</li> <li>*Reasons are clear but may be weak</li> <li>*Some irrelevant examples and quotes</li> <li>*Thesis is poorly written</li> </ul>	<ul style="list-style-type: none"> <li>*Questioned partly answered</li> <li>*Some reasons not clear</li> <li>*missing some examples and quotes</li> <li>*Thesis may be missing</li> </ul>	<ul style="list-style-type: none"> <li>*Off topic</li> <li>*Confusing reasons</li> <li>*Very few examples and quotes. Unclear.</li> </ul>
Vocabulary	<ul style="list-style-type: none"> <li>*Correctly used a <u>large</u> variety of words. (VTL, etc)</li> <li>*Very few spelling mistakes</li> </ul>	<ul style="list-style-type: none"> <li>*Correctly used a variety of words</li> <li>*Few spelling mistakes</li> </ul>	<ul style="list-style-type: none"> <li>*Correctly used a limited variety of words</li> <li>*Some spelling mistakes but meaning is clear</li> </ul>	<ul style="list-style-type: none"> <li>*Incorrect use of some words. Words too simple for level 3.</li> <li>*Many spelling mistakes</li> </ul>	<ul style="list-style-type: none"> <li>*Mostly incorrect use of words</li> <li>*Most spellings are wrong and confusing</li> </ul>
Organization, Cohesion, Style	<ul style="list-style-type: none"> <li>*Totally logical organization and very easy to read</li> <li>*Five well written paragraphs</li> <li>*Creative, well developed introduction, body and conclusion</li> <li>*A large variety of connecting words are used</li> </ul>	<ul style="list-style-type: none"> <li>*Logically organized and readable, but ideas may not connect smoothly</li> <li>*Four-Five paragraphs</li> <li>*Solid introduction, body and conclusion</li> <li>*A variety of connecting words are used</li> </ul>	<ul style="list-style-type: none"> <li>*Mostly readable with some confusing parts</li> <li>*Four paragraphs</li> <li>*Weak introduction, body and conclusion</li> <li>*Some connecting words are used</li> </ul>	<ul style="list-style-type: none"> <li>*Difficult to read. Not much organization of ideas.</li> <li>*Two-Three paragraphs</li> <li>*May be missing an introduction, conclusion, or body paragraph(s)</li> <li>*Missing connecting words, some are used incorrectly</li> </ul>	<ul style="list-style-type: none"> <li>*Very difficult to read</li> <li>*Text is not divided into paragraphs</li> <li>*Very few or no connecting words used</li> </ul>
Grammar and Form	<ul style="list-style-type: none"> <li>*A good range of simple, compound and complex sentences</li> <li>*Only one or two mistakes.</li> <li>*Excellent punctuation and capitalization</li> </ul>	<ul style="list-style-type: none"> <li>*Good use of simple sentences; some compound and complex sentences</li> <li>*Very few mistakes.</li> <li>*Very few punctuation and capitalization errors</li> </ul>	<ul style="list-style-type: none"> <li>*Mostly simple or compound sentences – a few complex attempted.</li> <li>*Some mistakes.</li> <li>*Some punctuation and capitalization errors.</li> </ul>	<ul style="list-style-type: none"> <li>*Mostly simple</li> <li>*Many mistakes but meaning can be found.</li> <li>*Many punctuation and capitalization errors.</li> </ul>	<ul style="list-style-type: none"> <li>*Sentence forms unclear</li> <li>*Difficult to understand- too many mistakes</li> <li>*Most punctuation and capitalization is incorrect.</li> </ul>

## Appendix 2

### Student and Teacher Project Scores

Student Scores		Teacher A	Teacher B
1.	12	12	12
2.	15	13.5	14
3.	16	15.5	16
4.	14	14	15
5.	18	16	18
6.	13	13	13
7.	14	14	14
8.	13	12	12.5
9.	16	16	16
10.	14	13.5	14
11.	15	14	14
12.	13	13	13
13.	14	14	14.5
14.	12	12	12.5
15.	19	18	18
16.	18	17	18
17.	16	16	16
18.	12	12	12
19.	14	14	14.5
20.	13	12	12.5
21.	15	15	16
22.	14	13	14
23.	16	15.5	15.5
24.	17	16	17
25.	12	12	12
26.	20	19	19
27.	15	15	15.5

28.	17	17	16
29.	16	15	16
30.	14	14	15
31.	13	12.5	13
32.	15	14	16
33.	16	16	16
34.	13	12	13
35.	16	15	16

### Appendix 3

#### Scatterplots



