# INTERNATIONAL JOURNAL of CONTEMPORARY EDUCATIONAL RESEARCH

Volume 2 | Issue 2 | Year 2015 | e-ISSN: 2148-3868

IJCER

# INTERNATIONAL JOURNAL of CONTEMPORARY EDUCATIONAL RESEARCH

IJCER

**Volume: 2 Issue: 2**

**June 2015**

## Editorial Board

## Indexing

**Directory of Open Access Journals**

**European Reference Index for the Humanities and Social Sciences (ERIH PLUS)**

**Scientific Indexing Services (SIS)**

**Türk Eğitim İndeksi (TEİ)**

**Open Academic Journals Index (OAJI)**

**Sosyal Bilimler Atıf Dizini (SOBİAD)**

## Table of Contents


**International Journal of Contemporary Educational Research (IJCER)**

[www.ijcer.net](http://www.ijcer.net)

# The Views of Pre-Service Teachers about the Use of Concept Cartoons in Science Courses

**Umit Izgi[1], Sevde Basar**
[1] Hacettepe University

**To cite this article:**

Izgi, U., & Basar, S. (2015). The views of pre-service teachers about the use of concept cartoons in science courses. *International Journal of Contemporary Educational Research, 2*(2), 61-68.

International Journal of Contemporary Educational Research
Volume 2, Number 2, 2015 July, Page 61-68 ISSN: 2148-3868

# The Views of Pre-Service Teachers about the Use of Concept Cartoons in Science Courses

**Umit Izgi**[1*], **Sevde Basar**
[1] Hacettepe University

## Abstract

The content of the science course include facts, concepts, principles and generalizations, theories and natural laws. Concepts refer to similar objects, human beings, events, views and processes. Concept learning is the basis for cognitive development. In recent years the following topics related to the science course have frequently been studied: conceptual development, teaching of concepts, concept maps, misconceptions, and conceptual change and concept cartoons. Concept cartoons were developed by Brenda Keogh and Stuart Naylor in the 1990s. Concept cartoons are visual teaching tools which allow for discussion of a scientific event through caricatures and which provides different perspectives about the event at hand. The aim of the study is to reveal the views of the pre-service science teachers about the use of concept cartoons in science teaching. The participants were senior pre-service teachers attending the science teaching division of Hacettepe University during the academic year of 2014-2015 and they all took the courses of special teaching methods I and special teaching methods II. The data of the study were collected through semi-structured interviews in which five open-ended items were asked to the participants. The data obtained were analyzed through content analysis method. Based on the findings of the study the effects of the use of the concept cartoons in science course on the teaching process were discussed based on the views of the pre-service teachers. Several suggestions were developed concerning the improvement of the efficacy of concept caricatures.

**Key words:** Science education, pre-service teachers, concept cartoons

## Introduction

In recent years it is widely agreed that traditional education cannot produce graduates with the desired qualification as a result of the advanced technology and accumulated knowledge. Therefore, those educational approaches which support active learners and learning by doing and which emphasize the higher thinking skills of individuals have begun to be used. In parallel to these developments the constructivist educational approach was introduced in Turkey in 2004. This approach argues that new knowledge can be based on previous learning and that learners should be active participants in teaching and learning process. It was employed in revision of the educational programs.

The teaching and learning process of the science course is regarded as significant. Because only through science knowledge individuals could make sense of environmental changes and developments and provide a logical framework for them. The science education program in Turkey was last developed in 2013. The ultimate goal of the program is to produce individuals who are science literate. Several qualities can be covered by the science literate. However, it mainly refers to those individuals who search for the changes in his environment, have higher levels of thinking and skills of scientific processes, are active participants of their learning, could make connections between their learning and daily life and are responsible for their learning. Various teaching and learning techniques are employed to make the individuals active in the teaching and learning process and to allow for them to acquire the desired knowledge and skills. One of such techniques is the concept cartoons.

Concept cartoons were developed by Brenda Keogh and Stuart Naylor in the 1900s and are based on the constructivist approach. These cartoons are designed to probe the views of the students about a science concept. In such cartoons the elements of humor are not used. Instead, these cartoons include different characters arguing about the answer to a question or debating alternate explanations of scientific phenomena. If there is a debate

* Corresponding Author: *Umit Izgi, umitizgi@gmail.com*

about alternate explanations of scientific phenomena, one of the arguments represents scientifically correct view and the others represent scientifically incorrect, but are held by students (Akamca and Hamurcu, 2009). Concept cartoons encourage students to think about a science concept and allow for them to actively take part in the learning process as well as provide a connection with the evaluation of learning (Keogh and Naylor, 2009). Students easily become interested in these cartoons which can be used as a teaching material.

Research suggests that concept cartoons are used as creative and complementary teaching tools in science education. The use of concept cartoons has become widely employed teaching and learning approach which may be used in different domains of learning (Dalacosta, Kamariotaki Paparrigopoulou, Palyvos and Spyrellis, 2009). In addition to be used for students, concept cartoons can also be used to improve the pre-service teachers' perspective regarding the teaching strategies that can be employed in science education. Because concept cartoons highlight the student misconceptions and reveal how they are active in questioning of their own ideas. Concept cartoons have been used as a complementary assessment technique in assisting pre-service teachers to control their science education perspective (Keogh and Naylor, 2000). Related studies show that concept cartoons are used to evaluate the comprehension levels of pre-service teachers about science concepts and topics in science and technology education (Chin and Teou, 2008; cited in: Dalacosta, Kamariotaki Paparrigopoulou, Palyvos and Spyrellis, 2009). On the other hand, it is argued that concept cartoons contribute to improve students' skills of problem-solving, critical thinking and producing scientific ideas and that these cartoons make the topic at hand much more interesting (Keogh and Naylor, 1999).

### Statement of the problem

The research question which the study tries to answer is as follows: What are the views of the pre-service science teachers who took the courses of special teaching methods I and special teaching methods II about the use of concept cartoons in science education?

Based on this research questions during the interviews the participants of the study were asked to answer the following five open-ended questions:

1. Do you think that the use of concept cartoons in science education contribute to student learning? Why?
2. At which stages can concept cartoons be used in science course? Please give details of this use.
3. Do you think that concept cartoons are much more influential in determining and reducing student misconceptions in science course in contrast to other teachniques such as concept maps, conceptual change sheets, etc? Why?
4. How can the efficacy of concept cartoons be improved?
5. What are the limitations in using concept cartoons in science course?

## Method

The study is designed as a case study. Case studies can be carried out using either quantitive or qualitative approach. Regardless of the approach adopted the goal of case studies is to describe a situation. In other words, case studies examine the related factors such as environment, individuals, events and processes about a situation using a holistic approach and focus on the effects of these factors on the case at hand. Given that each case has its own pecularities the results of the case studies cannot be generalized. On the other hand, these studies may provide necessary examples for other case studies which are carried out in similar situations (Şimşek and Yıldırım, 2006). In the study the data were collected through structured interviews, part of qualitative research technique.

### Participants

A total of 53 senior pre-service science teachers participated in the study. They were attending Hacettepe University during the academic year of 2014-2015. They took the courses of special teaching methods I and special teaching methods II in which they were studied and informed about concept cartoons.

### Data collection tools

The data of the study were collected through structured interviews in which participants were asked to answer five open-ended questions. The items were reviewed by a measurement-evaluation specialist and a field specialist in terms of content and coverage validity. Furthermore, the items were reviewed by a group of pre-service teachers who did not participated in the study in terms of understandability. Based on the feedbacks these items were finalized.

In scientific research validity and reliability are two significant points (Şimşek ve Yıldırım, 2000). In the study necessary analyses for validity and reliability were made. In terms of internal validity of the study it can be stated that the findings of the study were found to be significant in the framework of the study context and that these findings were interpreted based on this specific context. Consistency and significance of the findings were continuously checked. The concepts revealed were found to have common peculiarities. The findings obtained were found to be related to the conceptual and theoretical framework followed. As stated earlier, the generalization of the case study results is not so strong and therefore, their external validity is weak. However, the results could be extended to those case studies carried out in similar situations (Şimşek and Yıldırım, 2000). The findings of the present study could be tested in other case studies. In regard to its external reliability necessary explanations were provided about the data collection process, data processing process, the interpretation of the findings and the conclusions. The findings are related to the data obtained. It is stated that concerning internal reliability researcher should describe the research approach adopted and controls made at different phases of the study in a clear manner (Şimşek and Yıldırım, 2000). In regard to the internal reliability of the current study it can be stated that the research questions were explicitly given and explained. Therefore, the data obtained supported the findings of the study.

### Data analysis

Miles and Huberman (1994) argue that the process of data analysis is made up of three components: data reduction, data display and drawing conclusion and verification. During the data reduction researchers review and code the data they collected. In the coding process significant concepts and themes are used to choose those data which are significant.

The data then are displayed through graphics, tables and figures. For Miles and Huberman visualization of the data is very important in that it makes explicit the interrelations between concepts and between themes and it allows for reaching conclusions based on these concepts and themes.
In the final phase of the process, drawing conclusion and verification, the concepts, themes and relationships emerged are interpreted, contrasted and confirmed. Thus, the findings obtained become significant and valid.

The data collected were analysed following the data analysis process outlined by Miles and Huberman (1994). In the data analysis process the content analysis was employed and the frequency of the statements by the participants was found. In the discussion of these findings direct quotations were given to support the findings. Each quotation is followed by a code in brackets indicating the related pre-service teacher and the number given, such as (PSTX).

## Results

The findings obtained in the study are discussed in this section and are given in the following tables.

In regard to the first item asked in the structured interviews, "Do you think that the use of concept cartoons in science education contribute to student learning? Why?", Table 1 shows the views of the pre-service teachers about the contributions of concept cartoons in the teaching and learning process together with frequencies.

The answer of the participants to the question "Do you think that the use of concept cartoons in science education contribute to student learning? Why?" was positive and they all stated that the use of concept cartoons in science education contributes to student learning. Concerning the reasons for this contribution they mostly stated that concept cartoons provide long-lasting learning, avoid misconceptions and improve higher thinking skills of individuals. The following statement shows these ideas: *"Concept cartoons will allow children to learn and have fun in course. Moreover, these cartoons provide them with an opportunity to express their new ideas."* (PST 31); *"Concept cartoons reduce students' misconceptions and provide them with an opportunity to think critically. In addition, these concepts may provide an environment for productive discussions and for view exchange among students. If concept cartoons are used as an activity based on argumentation the higher thinking skills of students may be improved"* (PST 12); *"I think concept cartoons will contribute to the student learning. Through concept cartoons students learn with fun and with examples from daily life, making their learning long-lasting and significant ."* (PST 26).

**Table 1: Views of the participants about the contributions of concept cartoons in the teaching and learning process and frequency**

| Views | f |
|---|---|
| Yes, I do. Because | |
| Cartoons test the previous knowledge of the students | 2 |
| Cartoons make it easy to teach concepts | 3 |
| Cartoons provide effective teaching | 3 |
| Cartoons provides active student participation | 7 |
| **Cartoons provide long-lasting learning** | **17** |
| Cartoons provide feedback | 4 |
| Cartoons provide an opportunity to make connections with daily life. | 4 |
| Cartoons allow for courses to be smooth | 3 |
| Cartoons foster the learning process | 2 |
| **Cartoons reduces misconceptions of students** | **12** |
| Cartoons avoid the formation of misconceptions. | 1 |
| **Cartoons improve higher thinking skills.** | **11** |
| Cartoons improve communication skills. | 1 |
| Cartoons direct students to make search. | 1 |
| Cartoons provide an environment eligible for discussion. | 7 |
| Cartoons enrich the learning environment. | 1 |

In regard to the second item asked in the structured interviews, "At which stages can concept cartoons be used in science course?", Table 2 shows the views of the pre-service teachers about the use of concept cartoons at different stages of science course together with frequencies.

**Table 2: Views of the participants about the use of concept cartoons at different stages of science course**

| Views | f |
|---|---|
| **a . Introduction** | **34** |
| a.1. Getting attention | 21 |
| a.2. Improving motivation | 21 |
| a.3. Measuring readiness of students | 7 |
| **b. Lecture** | **34** |
| b.1. Concept teaching | 5 |
| b.2. Expansion of topics | 5 |
| b.3. Briefing | 3 |
| b.4. Repetition of topics studied | 2 |
| b.5. Correction of misconceptions | 9 |
| **c. End of course** | **35** |
| c.1. Inspection of incomplete knowledge | 1 |
| c.2. Measuring and evaluating the achievement of the goals of the course | 17 |

In response to the question "At which stages can concept cartoons be used in science course?" the participants stated that concept cartoons can be used at introduction (34 participants), lecture stages (34 participants) as well as at the end of the course (35 participants). Their views are as follows:
*"Concept cartoons can be used at the introduction part to improve student motivation. Using the cartoons at this stage will improve student attention and help them to focus on the course.*" (PST 27); *"For me concept cartoons can be used at any stage of the course. If these are used at the introduction part they will improve student attention; if it is used at the evaluation part it may indicate at which level the course goals are achieved."* (PST 2); *"Concept cartoons can be used at the introcduction stage to make it possible to have better understanding of science concept and to allow for students to think about it and the cartoons can be used at the end of the course to consolidate student learning."* (PST 7).

In regard to the third item asked in the structured interviews, "Do you think that concept cartoons are much more influential in determining and reducing student misconceptions in science course in contrast to other teachniques such as concept maps, conceptual change sheets, etc? Why?", Table 2 shows the views of the pre-service teachers about the effects of the use of concept cartoons on student misconceptions with frequencies.

**Table 3: Views of the participants about the use of concept cartoons in reducing student misconceptions**

| Views | f |
|---|---|
| a. No | 9 |
| a.1. Concept maps are much better for this aim. | 1 |
| a.2. Conceptua change sheets are much better for this aim. | 4 |
| a.3. Each method is significant and has its own specific usage. | 9 |
| b. **Yes** | **29** |
| b.1. Concept cartoons are better than concept maps in reducing student misconceptions | 2 |
| b.2. Concept cartoons are better than concept change sheet in reducing student misconceptions | 2 |
| c. Undecided | 3 |

In response to the question "Do you think that concept cartoons are much more influential in determining and reducing student misconceptions in science course in contrast to other techniques such as concept maps, conceptual change sheets, etc.? Why?", nine participants argued that in reducing student misconceptions concept cartoons are not more influential than other methods. One of these participants claimed that concept maps are the best method of reducing student misconceptions and four considered conceptual change sheets to be the best method of reducing student misconceptions. Nine other participants stated that each method has its own use and significance in this regard. The number of the participants who regarded concept cartoons as the best method of reducing student misconceptions is twenty-nine. Three participants did not provide an answer to this question.

The statements of the participants concerning the use of concept caricatures and other methods in reducing misconcepts are as follows: *"For me concept maps should be used to teach more general topics at the beginning or end of the course, but concept cartoons should used for the concepts which are significant and can be confused. I think concept cartoons are much more special method."* (PST 14); *"I think that concept cartoons are not superior than other methods in reducing misconceptions. All methods may facilitate learning and teaching as long as teacher is aware of which method should be used where."* (PST 19), "*I think that it is not possible to argue any method is superior than the others. Each has its own usage. However, I myself prefer to use concept cartoons in reducing student misconceptions."* (PST 9).

In regard to the fourth item asked in the structured interviews, "How can the efficacy of concept cartoons be improved?", Table 4 shows the views of the pre-service teachers about the ways to improve efficiency of concept cartoons.

**Table 4: Views of the pre-service teachers about the ways to improve efficiency of concept cartoons**

| Views | f |
|---|---|
| **About teachers and student teachers** | |
| **They may be taught about concept cartoons.** | **9** |
| The ministry may deal with it. | 1 |
| A course on concept cartoons may be covered in teacher training programs. | 2 |
| More cocnept cartoons may be used in courses | 5 |
| Collaborations may be developed with professional cartoonists. | 1 |
| **About concept cartoons** | |
| Cartoons may include more characters. | 2 |
| Cartoons may involve humor. | 3 |
| Cartoons may include more misconceptions. | 4 |
| Cartoons may include empty blank speech bubbles to be filled by students | 3 |
| Characters may be named or numbered | 1 |
| Cartoons may use shorter sentences | 7 |
| Cartoons may include more vivid colors | 2 |
| Cartoons may be developed for different age groups. | 9 |
| Cartoons may include those questions to ask students to think or discuss about | 7 |
| Cartoons may include topics and characters from daily life. | 2 |
| **Cartoons may be about topics and gains** | **12** |
| **About teaching process** | |
| **Cartoons may be combined with such techniques as argumentation** | **3** |
| **About student** | |
| **Students may prepare concept cartoons** | **2** |
| Students may be offered a course on concept cartoons. | 1 |

The answers to the question of "How can the efficacy of concept cartoons be improved?" showed that the participants suggested four different ways to improve the efficacy of concept cartoons. More specifically, nine participants argued that teachers and student teachers may be taught about the use of concept cartoons. There were also twelve participants who suggested that concept cartoons should be about topics and gains. The combination of concept cartoons with other techniques such as argumentation was suggested by three participants. In order to improve the efficacy of the concept cartoons two participants argued that students should be involved in the development process of concept caricatures.

The views of the participants are exemplified as follows: *"Teachers should be informed about concept cartoons which may be involved in in-service activities. Concept cartoons should involve those pictures and dialogues in pertinent to the topic at hand."* (PST 34), *"Concept cartoons should be more than drawings and focus on the targeted topic or concept. Cartoons should make active student involvement possible and provide a discussion environment. Cartoons should also be clear and easy to understand and should be about the gains."* (PST 38); "*The topics at hand should be given in a simplified form. Each cartoon should deal with a single topic. Interesting visuals should be employed instead of those which are unnecessary. Any expression which may cause misconception should be avoided. Cartoons should be consistent with the topics to be studied."* (PST 39).

Table 5 shows the views of the participants about the limitations in using concept cartoons in science courses.

**Table 5: Views of the participants about the limitations in using concept cartoons in science courses**

| Views | f |
|---|---|
| 1. There is no limitation in using concept cartoons. | 2 |
| 2.There are some limitations | 14 |
| 2.1. Cartoons may not be used for each phase of the course. | 1 |
| 2.2. Use of cartoons may take longer time. | 2 |
| 2.3. Cartoons may be ineffective in explaining the concepts. | 1 |
| 2.4. Concept cartoons may lead to distraction. | 3 |
| 2.5. Students may focus on characters instead of topic. | 1 |
| 2.6. For each topic cartoons may not be developed. | 7 |
| 2.7. Cartoons may not allow to teach topic as a whole. | 5 |
| 2.8. Cartoons may make it difficult to manage the class for teacher. | 1 |
| 3. Undecided | 2 |

In response to the question of "What are the limitations in using concept cartoons in science course?" sixteen participants argued that there is no such limitation. The number of the participants who thought that there are some limitations was fourteen. These participants argued that cartoons may not use at every step of the course and that the use of concept cartoons make take longer time and lead to insufficient explanation of concepts, to distraction and to focus on characters instead of the topic. They also stated that cartoons may not be developed for each topic and may lead to difficulties in the management of classroom. Two participants did not give any comment on the limitations about the use of concept cartoons.

The following statements exemplify the views of the participants about the limitations about the use of concept cartoons: *"I think there is no limitation in using concept cartoons. Teachers may easily prepare and use them in the course. It is economic and does not lead to extracurricular assignment."*; "*I think there is no limitation in using concept cartoons. Because concept cartoons aim to provide the key concepts in the course and to integrate the previous and future learning. I think cartoons serve for these goals."* (PST 40), *"Cartoons may not be used at each and every step of lecturing. It is not possible to have full understanding of any topic using concept cartoons. Therefore, cartoons may not be used for each topic to be studied."* (PST 5).

## Discussion and Conclusion

The findings of the current study are mostly consistent with the previous findings. All of the participants stated that concept cartoons positively affect the learning process. Concerning the reasons for this contribution they mostly stated that concept cartoons provide long-lasting learning, avoid misconceptions and improve higher thinking skills of individuals. Similarly, Keogh and Naylor (1999) found that both pre-service teachers and students had positive views about the use of concept cartoons. Also in the study of Şaşmaz Ören and Meriç (2014), the most important finding from students' journals is that students have used a large number of positive sentences for this technique. Students have developed a positive attitude towards Science courses. İzgi and Kaptan (2010) found that 20.1% of the pre-service teachers regarded the science and technology course as the best eligible course for the use of concept cartoons. It was also found that 15% of the participants considered concept cartoons as an enjoyable method and that 13.44% of them regarded concept cartoons as an effective method in determining and reducing students' misconceptions. Düzgün (2013) also concluded that the pre-service teachers considered concept cartoons as a useful, visual, interesting and multi-purpose tool.

In response to the question "At which stages can concept cartoons be used in science course?" the participants stated that concept cartoons can be used at introduction (34 participants), lecture stages (34 participants) as well as at the end of the course (35 participants). İzgi and Kaptan (2010) concluded that 25.3% of the pre-service teachers regarded it as proper for the evaluation of student learning.

In response to the question "Do you think that concept cartoons are much more influential in determining and reducing student misconceptions in science course in contrast to other techniques such as concept maps, conceptual change sheets, etc.? Why?", nine participants argued that in reducing student misconceptions concept cartoons are not more influential than other methods. One of these participants claimed that concept maps are the best method of reducing student misconceptions and four considered conceptual change sheets to be the best method of reducing student misconceptions. Nine other participants stated that each method has its own use and significance in this regard. The number of the participants who regarded concept cartoons as the best method of reducing student misconceptions is twenty-nine. Three participants did not provide an answer to this question. Ekici, Ekici and Aydın (2007), analyzed the role of concept cartoons in determining and reducing student misconceptions about photosynthesis based on student views. They concluded that concept cartoons are not effective only in determining the student misconceptions about photosynthesis, but also in reducing these misconceptions.

The answers to the question of "How can the efficacy of concept cartoons be improved?" showed that the participants suggested four different ways to improve the efficacy of concept cartoons. More specifically, nine participants argued that teachers and student teachers may be taught about the use of concept cartoons. There were also twelve participants who suggested that concept cartoons should be about topics and gains. The combination of concept cartoons with other techniques such as argumentation was suggested by three participants. In order to improve the efficacy of the concept cartoons two participants argued that students should be involved in the development process of concept caricatures.

In response to the question of "What are the limitations in using concept cartoons in science course? " sixteen participants argued that there is no such limitation. The number of the participants who thought that there are some limitations was fourteen. These participants argued that cartoons may not use at every step of the course and that the use of concept cartoons make take longer time and lead to insufficient explanation of concepts, to distraction and to focus on characters instead of the topic. They also stated that cartoons may not be developed for each topic and may lead to difficulties in the management of classroom. Two participants did not give any comment on the limitations about the use of concept cartoons. In the study by İzgi and Kaptan (2010) the pre-service teachers participated in the study reported that cartoons may facilitate discussions in class through characters with speech bubbles and may be used to determine and reduce misconceptions at every step during the course. They also stated that cartoons may reduce the negative attitudes of the students towards science courses due to the elimination of conceptual confusion. In addition, they argued that concept cartoons improve student attention towards the science and technology course and therefore, can be used in the science and technology courses.

Based on the findings of the study about the use of concept cartoons based on the views of the pre-service teachers and previous findings the following points are suggested:

- In order to improve the efficacy of concept cartoons pre-service teachers should be fully informed about these cartoons and in-service teachers should be informed about them through in-service training activities.
- Written teaching materials, especially textbooks, should contain more information about the technique of concept cartoons.
- The limitations of the use of concept cartoons in other courses should be identified in order to minimize them.

## References

Akamca, Özyılmaz G. ve Hamurcu, H. (2009). Analojiler, kavram karikatürleri ve tahmin-gözlem-açıklama teknikleriyle desteklenmiş fen ve teknoloji eğitimi. *E-journal of New World Sciences Academy,* volume: 4, number: 4, article number: 1C0089.

Dalacosta, K.; Kamariotaki Paparrigopoulou, M.; Palyvos, J.A. ve Spyrellis, N. (2009). Multimedia application with animated cartoons for teaching science in elementary education. *Computers & Education* 52, 741–748.

Düzgün, E. M. (2013). *Sınıf öğretmeni adaylarının fen ve teknoloji dersinde kullanılan kavram karikatürlerine yönelik görüşleri.* Yayımlanmamış Yüksek lisans tezi. Afyon Kocatepe Üniversitesi, Sosyal Bilimler Enstitüsü, Afyon.

Ekici, F., Ekici, E., & Aydın, F. (2007). Utility of Concept Cartoons in Diagnosing and Overcoming Misconceptions Related to Photosynthesis. *International of Journal of Environmental & Science Education, 2*(4), 111-124.

İzgi, Ü. ve Kaptan, F. (2010). Fen ve teknoloji öğretiminde öğretmen adaylarının kavram karikatürlerini kullanımı üzerine bir çalışma. *Uluslararası Öğretmen Yetiştirme Politikaları ve Sorunları Sempozyumu II*, Hacettepe Üniversitesi, Ankara.

Keogh, B. ve Naylor, S. (1999). Concept cartoons, teaching and learning in science: an evaluation. *International Journal Of Science Education.*21(4). 431- 446.

Keogh, B. ve Naylor, S. (2000). Teaching & learning in science using concept cartoons: why Dennis wants to stay in at playtime. *Australian Primary & Junior Science Journal;* August, vol. 16, issue 3, p10, 5p, 2bw.

Keogh, B. ve Naylor, S. (2009). Active assessment. *Mathematics Teaching*, 215,35–37, September.

Miles, M. B. Ve Huberman, M. (1994). *Qualitative data analysis: an expanded sourcebook* (2. baskı). Thousand Oaks, CA: Sage.

Sasmaz Oren, F. & Meric, G. (2014). Seventh grade students' perceptions of using concept cartoons in science and technology course. *International Journal of Education in Mathematics, Science and Technology*, 2(2), 116-137.

Şimşek, H. ve Yıldırım, A. (2006). *Nitel araştırma yöntemleri*. Ankara: Seçkin Yayıncılık.


International Journal of Contemporary Educational Research (IJCER)

www.ijcer.net

# The Effects of STEM PBL on Students' Mathematical and Scientific Vocabulary Knowledge

**Ali Bicer[1], Peter Boedeker[1], Robert M. Capraro[1], Mary M. Capraro[1]**
[1] Texas A&M University

**To cite this article:**

Bicer, A., Boedeker, P., Capraro, R.M., & Capraro, M.M. (2015). The effects of STEM PBL on students' mathematical and scientific vocabulary knowledge. *International Journal of Contemporary Educational Research, 2*(2), 69-75

**International Journal of Contemporary Educational Research**
Volume 2, Number 2, July 2015, Page 69-75 ISSN: 2148-3868

# The Effects of STEM PBL on Students' Mathematical and Scientific Vocabulary Knowledge

**Ali Bicer[1*], Peter Boedeker[1], Robert M. Capraro[1], Mary M. Capraro[1]**

[1] Texas A&M University

## Abstract

Vocabulary is at the surface level of language usage; thus, students need to develop mathematical and scientific vocabulary to be able to explicitly communicate their mathematical and scientific reasoning with others. The National Council of Teachers of Mathematics (NCTM) and the National Science Teachers Association (NSTA) have both created communication standards within mathematics and science disciplines. In the present study, science, technology, engineering, and mathematics (STEM) project based learning (PBL) methods were utilized during a summer camp in 2013 to encourage interest in and grow the knowledge of students in the STEM disciplines. The participants ($N$ = 53; 18 female, 35 male, 5 Asian, 6 African American, 12 White, and 30 Hispanic) were 8$^{th}$ grade students. The paired-sample $t$ tests' results showed that the model of STEM PBL instruction elicited a statistically significant ($p < 0.05$) improvement in the mathematical and scientific vocabulary knowledge of students with the *Cohen's d* effect size of 0.62 and 0.84 respectively. STEM PBL could be a beneficial instructional method concerning vocabulary mastery for students in science and mathematics classes.

**Key words:** STEM PBL, Mathematics and Science vocabulary.

## Introduction

Every day people communicate using words and expressions that may have various meanings. In order to fluently communicate with one another, people must have a shared understanding of the meaning of the words they use. Otherwise, the communication could fall short of conveying the intended message. Individuals who spoke entirely different languages could have great difficulty verbally communicating because of the differences in words used for the same object. Clarity in language facilitates communication and could be the building block to greater understanding, particularly in the fields of science and mathematics.

Developing vocabulary in the science and mathematics classrooms was much like learning a foreign language. When a new word was introduced to a student, the word had to be taught and the meaning explained. Once this was achieved, students could communicate meaningfully with their peers and teachers. The development of a common vocabulary within a subject was vital so that individuals could communicate coherently and efficiently with one another. Development of vocabulary in the content areas of mathematics and science was important for students to be able to communicate and more completely understand the topics they were learning. Vocabulary was often built on previous vocabulary, with one word perhaps representing a single process that required understanding of several other content specific vocabulary words. For instance, the word "photosynthesis" represented the process of a plant producing energy from the sun, but within that process words such as "chlorophyll," "organelles," and "chloroplasts" needed to be understood. Determining the best methods of teaching vocabulary to students was critical.

In 2003, the National Council of Teachers of Mathematics (NCTM) emphasized the importance of language in mathematics classrooms by placing a broad range of mathematical communication goals into NCTM's recommended standards. These goals included, but were not limited to, students being able to: a) organize their mathematical thinking, b) communicate in the mathematics classroom with others (teachers and peers), c) evaluate different mathematical solution strategies, and d) use mathematical language explicitly. However,

[*] Corresponding Author: *Ali Bicer, alibicer@email.tamu.edu*

achieving these goals in mathematics classrooms was not an easy job. Thompson and Rubenstein (2000) noted that using mathematical language was often a challenge for students due to several reasons. One of the common reasons was that students lacked mathematical vocabulary development. Previous studies found that students have difficulty communicating their own mathematical thoughts and understanding the meaning of mathematical texts due to their lack of mathematical vocabulary development (Rubenstein, 2007; Kenney et al., 2005). According to Miller (1993), "without an understanding of the vocabulary that is used routinely in mathematics instruction, textbooks, and word problems, students are handicapped in their efforts to learn mathematics" (p. 312). Different approaches had been taken to determine the best method of increasing student mathematical vocabulary.

There were several challenges to learning mathematics vocabulary. When students were able to identify a relationship between new vocabulary words and their prior knowledge, the process of learning new vocabulary was easier and avoided more misconceptions (Ashlock, 2006). However, Cirillo, Bruna, and Herbal-Eisenmann (2010) noted that the mathematical vocabulary students learned during instruction was mostly limited to the classroom and students did not have an opportunity to apply the words in their daily lives. The inability to apply the mathematical terminology in a real world setting inherently limited the opportunities for learning reinforcement. In addition to the lack of real world application, there were several other challenges students had to overcome in order to develop a fluent and accurate understanding of mathematics vocabulary, including: a) some mathematical words were shared with everyday English but sometimes with distinct meanings, b) some mathematical terms were used only in mathematics, c) some mathematical words had multiple meaning depending on context, d) some mathematical words had the same sound as everyday English words, e) some words had modifiers that may change their meaning, f) some words were also used in science, and g) some words were learned in pairs (Rubenstein, 2013; Thompson & Rubenstein, 2000). Teachers in the mathematics classroom should allow students to explore, investigate, and explain mathematical vocabulary when they encounter new mathematical words and terms (Steele, 1999).

**Science Vocabulary**

Vocabulary mastery was also important to success in the science classroom. Learning science included learning the proper terminology or vocabulary for the science discipline being taught. The National Science Teachers Association (NSTA, 2014) developed the Next Generation Science Standards in which students were required to communicate design ideas, solutions, and scientific and technical information through oral presentations or written forms. These different means of communicating in a professional and meaningful way were to develop the skills that scientists and engineers would need in their careers. These methods of communication also showed greater fluency with new, content-specific vocabulary than simply filling in the blank or matching tasks. A questionnaire given by Cook and Tulip (1992) to science teachers to determine what features science teachers felt were important in a textbook found that a glossary (vocabulary) ranked 12$^{th}$ out of 40 with a mean of 4.46 (features were ranked from 1 to 5 with 5 being "very important"). Science vocabulary mastery was important to developing proficiency in the sciences.

Science vocabulary, much like the mathematics vocabulary previously discussed, was both critical and challenging for a student to master. One of the challenges associated with learning science vocabulary was the fact that comprehension of a new word depended greatly on understanding a variety of other vocabulary words (Fisher, Grant, & Frey, 2009). Within a single word there may be "stored descriptions and explanations of ideas, events, or patterns" (Yore, Craig, & Maguire, 1998, p. 34). A student that struggled with basic science vocabulary found the mastery of more advanced topics difficult because mastery required synthesizing multiple science vocabulary words to understand complex systems and processes (Cohen, 2012). For instance, a student that did not comprehend the meaning behind the word "evaporation" would not be able to fully understand the complex system of the water cycle, of which "evaporation" is only a part. In addition to the compounding nature of science vocabulary, the extensive number of science vocabulary words presented a challenge. Within textbooks there could exist hundreds or even thousands of new science words (Groves, 1995). This large number of new science vocabulary that existed in textbooks could present students with an overwhelming task as they tried to master all of them. The fact that science textbooks presented new vocabulary words and then used those very words to present even greater concepts, coupled with the extensive nature of science vocabulary lists, made mastery of science words critical and challenging for students.

***Methods of Teaching Vocabulary***

To enable students to overcome the obstacles to learning new, content-specific vocabulary, researchers have suggested several instructional models for mathematics and science. For this study, two broad categories had been identified for methods of instruction to increase science and mathematics vocabulary. The first method was through contextual learning, and the second concentrated on more direct methods.

Contextual instructional methods enabled students to witness the processes and actions behind the vocabulary, allowing students to create a mental image. NCTM (1989; 1991) and Miller and Gildea (1987) encouraged using a contextual mathematical instructional model, thus allowing students to observe how mathematical words were used in a mathematical context. Contextual learning could help students to develop images of the word meanings, thereby creating a deeper understanding of the term. Several different imagery-based interventions could be utilized to help students form mental images of the science concepts being taught, thus making the meanings of the associated vocabulary words more easily retained and recalled (Cohen & Johnson, 2011; Cohen, 2012).

Direct teaching methods focused on in-class activity that was less hands-on than contextual learning but still required students to engage in learning in non-traditional ways. Vacca and Vacca (1996) noted that important mathematical words needed to be taught by direct instruction. Methods that have been employed to teach science vocabulary have included using text cards, word lists, graphic organizers, and word games (Carrier, 2011). Developing a strong vocabulary foundation at a young age and continuing to foster new vocabulary mastery through literacy-based interventions could be beneficial in learning new vocabulary (Cohen, 2012).

Monroe and Orme (2002) found that neither contextual nor direct vocabulary teaching alone was sufficient to develop students' mathematical vocabulary, but these two instructional methods should be complementary to each other. Within a lesson, vocabulary must be thoughtfully integrated in such a manner that conceptual learning was not trumped by simple memorization of new terminology (Bay-Williams & Livers, 2009) utilizing the appropriate methods available. This perspective, that the best method of teaching vocabulary was the combination of direct and contextual learning, was the theoretical basis of the present study.

The cohesion of findings in previous research provided the framework for the present study. Learning vocabulary beyond the rote memorization of terms and definitions required a combination of direct teaching in which an instructor gave the meanings of words and allowed students the opportunity to develop an understanding of the processes and purposes that each vocabulary word represents (Monroe & Orme, 2002). By working in a context, students were able to associate images and actions with terms and reinforce their understanding of definitions. Instructional methods that provide students with the opportunity to learn definitions of words associated with a context and develop imagery related to each term would benefit students' vocabulary retention and understanding. Because project based learning (PBL) includes features of both contextual and direct teaching and allows students to create mental images associated with scientific vocabulary, the present study attempted to determine how STEM PBL (see the conceptual framework in Figure 1) affected students' mathematical and scientific vocabulary knowledge.


Figure 1. *Conceptual Framework*

### *STEM Project Based Learning (STEM PBL)*

Understanding how students learn best could aid the teacher in determining how to teach vocabulary words. Project based learning was an instructional method driven by student inquiry and directed by teacher guidance (Bell, 2010). Projects were created by students and shared with their peers, and outcomes of PBL implementation included greater and deeper understanding of topics, higher-level reading, and an increase in motivation. PBL challenged students with authentic tasks in real world contexts to develop and use their knowledge of different subjects (Thomas, 2000). Science, technology, engineering, and mathematics (STEM) PBL was the application of PBL methods in one of the STEM subjects or in an interdisciplinary manner (Corlu, Capraro, & Capraro, 2014). Introducing new vocabulary words through STEM PBL helped students to connect new vocabulary words with imagery in a concrete way (Bicer, Navruz, Capraro, & Capraro, 2014; Bicer, et al., 2015). When a student was involved with the concept in a hands-on way and used that new vocabulary word in action, the student was better prepared to write technical information using proper terminology or vocabulary (Bicer, Capraro, & Capraro, 2013; Bicer, Capraro, & Capraro, 2014). The use of PBL for increasing vocabulary mastery in both mathematics and science could be promising due to the contextual application of the hands-on activity and the potential for students to develop imagery connecting vocabulary words to the meaningful actions and processes taking place, as well as by providing opportunities for teachers to give direction and guidance.
The present study aimed to answer the following question:

(1) How does engagement in a STEM PBL activity affect students' mathematical and scientific vocabulary development?
(2) How does engagement in a STEM PBL activity affect students' scientific conceptual understanding?

## Method

### Participants

The participants ($N$ = 53) were 8th grade students who attended a summer camp in 2013. Of the 53 students, 18 were female and 35 were male. The student group was comprised of 5 Asian, 6 African American, 12 White, and 30 Hispanic students.

### Intervention

The intervention took place during a two-week summer camp designed by a STEM center at a Tier 1 research university. Throughout the summer camp, students were involved in activities that fostered their mathematical and scientific vocabulary knowledge through STEM PBL. One of the projects students engaged in during the camp was an egg drop PBL that consisted of designing and testing a unique parachute that would land an egg safely on the ground. These parachute activities met four days per week for one hour and 15 minutes per day and included 10 hours per week of independent study including, but not limited to, gathering materials, creating hypotheses, and journal writing.

### *Vocabulary Assessment*

The study involved 24 academic vocabulary words, of which 12 were mathematics words and 12 were science words. Students learned and repeatedly used these terms during the egg drop PBL activity. The mathematical vocabulary words and terms were: pi, Pythagorean Theorem, constant, regular polygon, diameter, variable, angle, hexagon, surface area, polygon, circle, and length segment. These mathematical vocabularies were selected as being more problematic from the list provided by Rubenstein (2013) and Thompson and Rubenstein (2000). The science vocabulary words and terms were: velocity, gravity, energy, force, kinetic energy, potential energy, momentum, air resistance, thermal energy, terminal velocity, friction, and the law of conservation. These science vocabularies were selected based on the science standards from the Texas Education Agency (TEA). Students were required to use the selected words to effectively communicate their mathematics and science reasoning during the PBL with their peers and teachers. Pre and post-test examinations were administered to assess student vocabulary knowledge of the words that were emphasized during the PBL. These exams consisted of definitions and a word bank of vocabulary words from which to select the term that appropriately matched each definition. Each correct match was assessed as a score of one point with a possible total score of 12 points per subject. Another test that included true-false statements was administered in order to assess whether students increased their scientific conceptual understanding. This test (see Table 1) was utilized to determine if learning through a STEM PBL activity also had an effect on students' conceptual understanding. Each correct answer was assessed as a score of 1 point with a possible total of 11 points.

Table 1. True-False Statements Test

| | |
|---|---|
| 1) T-F | Energy is greatest in the egg drop vehicle just before it hits the ground. |
| 2) T-F | Before the egg starts to fall all of its energy is in the form of kinetic energy. |
| 3) T-F | Momentum decreases as the egg falls. |
| 4) T-F | Momentum depends on both the speed of the object and its mass. |
| 5) T-F | The force of gravity acting on the egg is greater than the force of air resistance. |
| 6) T-F | The force of gravity gets larger the more mass it has. |
| 7) T-F | The air resistance on an object depends on the mass of the object. |
| 8) T-F | Gravity and Momentum both stay the same the whole time the egg falls. |
| 9) T-F | Potential energy increases the closer an object is to the ground. |
| 10) T-F | Kinetic Energy increases as the speed increases. |
| 11) T-F | Kinetic energy starts out at 0. |

### *Analysis*

The Statistical Package for Social Science (SPSS) version 17.0 (2008) was used to run the analysis on 53 written responses. There were no missing data, and this sample size was adequate to detect differences between two dependent means (paired *t* tests) at the 5% statistical significance level (Faul, Erdfelder, Lang, & Buchner, 2007). Cohen's d effect sizes and confidence intervals were provided to examine the practical importance of the present study.

## Results

The paired-sample *t* test results showed that STEM PBL instruction elicited a statistically significant improvement in the mathematical vocabulary knowledge of students, $p < 0.05$ with the effect size of *Cohen's d* = 0.62. Results also showed that the model of STEM PBL instruction statistically significantly increased students' science vocabulary knowledge, $p < 0.05$, *Cohen's d* = 0.84. Students' pre-science and pre-mathematics scores were reexamined to eliminate the scores already perfect in the pre-test. The paired-sample t tests were also used to test how students' scientific conceptual understandings changed after participating in STEM PBL. The results associated with the true-false science concepts test indicated that students increased their mean score of scientific conceptual understanding; however, this increase (see Figure 1) on their posttest compared to their pretest was not statistically significant, $p > 0.05$, *Cohen's d* = 0.14.

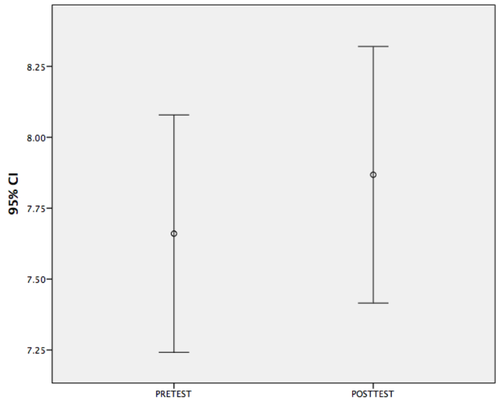



Figure 1. *A 95% Confidence Interval for Students Scientific Conceptual Understanding*

## Discussion

Vocabulary development begins at a young age and continues throughout a person's life. A single word can indicate various and complex meanings. This reality makes vocabulary mastery critical, especially when students begin to study more complex mathematics and science topics. A word may indicate an entire process or theorem and have differing meanings depending on the context. Shortcomings in these specialized vocabularies lead to challenges when trying to communicate with peers and teachers (Miller, 1993; Rubenstein, 2007; Kenney et al., 2005). NCTM (2003) and NSTA (2014) have called for an increase in communication skills within mathematics and science studies. Ensuring that students are able to share and explain ideas and solutions fluently with peers and teachers could be facilitated by improved student vocabulary in the content areas of mathematics and science.

The purpose of this study was to test whether participating in STEM PBL activities affects students' mathematics and science vocabulary knowledge. STEM PBL integrates direct teaching with contextual learning and allows students to develop mental images and make connections between content-specific vocabulary and activities taking place in the lab or classroom. An important aspect of STEM PBL is that students share their solutions with their peers, thus strengthening their ability to communicate using terms that are content specific (Thomas, 2000). As these communication skills are developed, the vocabulary is internalized through the hands-on activities and interactions of the students in their classes. Students are given the opportunity to see visuals and create mental images of scientific processes associated with vocabulary words. These images can help students retain vocabulary meanings and enhance students' recall of definitions (Cohen, 2012).

The present study indicates favorable outcomes when implementing STEM PBL. More can be done to determine the value of STEM PBL implementation. Studies evaluating the effectiveness of STEM PBL during the regular school day would be informative, considering the fact that this particular study took place during a summer camp. Research with a larger group of students would also yield more interesting results. Further studies can be conducted to give greater insight into the benefits of STEM PBL as a means of vocabulary and conceptual learning.

## References

Ashlock, B. R. (2005). Error patterns in computation: Using error patterns to improve instruction. Upper Saddle River, NJ: Prentice Hall.

Bay-Williams, J. M., & Livers, S. (2009, November). Supporting math vocabulary acquisition. Teaching ChildrenMathematics, 16(4), 238-245.

Bell, S. (2010). Project-based learning for the 21st century: Skills for the future. The Clearing House, 83, 39-43.

Bicer, A., Capraro, R. M., & Capraro, M. M. (2013). Integrating writing into mathematics classrooms to increase students' problem solving skills. International Online Journal of Educational Science, 5(2), 361-369.

Bicer, A., Navruz, B., Capraro, R. M., Capraro, M.M., Oner, T.A., & Boedeker, P. (2015). STEM schools vs. non-STEM schools: Comparing students' mathematics growth rate on high-stakes test performance. International Journal of New Trends in Education and Their Implications, 6(1), 138-150.

Bicer, A., Navruz, B., Capraro, R. M., & Capraro, M. M. (2014). STEM schools vs. non-STEM schools: Comparing students' mathematics state based test performance. International Journal of Global Education, 3(3), 8-18.

Bicer, A., Capraro, R. M., & Capraro, M. M. (2014). Integrating writing into mathematics classroom as one communication factor. *The Online Journal of New Horizon in Education*, *4*(2), 58-67.

Carrier, S. J. (2011). Effective strategies for teaching science vocabulary. Retrieved from http://www.learnnc.org/lp/pages/7079

Cohen, M. T., & Johnson, H. L. (2011). Improving the acquisition of novel vocabulary through the use of imagery interventions. Early Childhood Education Journal, 38, 357-366.

Cohen, M. T. (2012). The importance of vocabulary for science learning. Kappa Delta Pi Record, 48(2), 72-77.

Cook, A. & Tulip, D (1992). The Importance of Selected Textbook Features to Science Teachers. Research in Science Education, 22, 91-100. Retrieved from http://link.springer.com/article/10.1007/BF02356883

Corlu, M. S., Capraro, R. M., & Capraro, M. M. (2014). Introducing STEM education: Implications for educating our teachers in the age of innovation. Education and Science, 39(171), 74-85.

Cirillo, M., Bruna, R. K., & Herbel-Eisenmann, B. (2010). Acquisition of mathematical language: Suggestions and activities for English language learners. Multicultural Perspectives, 12 (1), 34-41.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis for the social, behavioral, and biomedical sciences. Behavior Research Methods, 39, 175-191.

Fisher, D., Grant, M., & Frey, N. (2009). Science literacy is greater than strategies. Clearing House: A Journal of Educational Strategies, Issues and Ideas, 82(4), 183-86.

Groves, F.H. (1995). Science vocabulary load of selected secondary science textbooks. School Science and Mathematics, 95(5), 231-235.

Kenney, J. M., Hancewicz, E., Heuer, L., Metsisto, D., & Tutle, C. L. (2005). Literacy strategies for improving mathematics instruction. Alexandra, VA: Association for Supervision and Curriculum Development.

Monroe, E., & Orme, M. (2002). Developing mathematical vocabulary. Preventing School Failure, 46(3).

Miller, D. L. (1993). Making the connection with language. Arithmetic Teacher, 40(6), 311-316.

Miller, G. A., & Gildea, P. M. (1987). How children learn words. Scientific American, 257, 94-99.

National Council of Teachers of Mathematics. (1991). Professional standards for teaching mathematics. Reston, VA: Author.

National Council of Teachers of Mathematics. (1989). Curriculum and evaluation standards for school mathematics. Reston, VA: Author.

National Council of Teachers of Mathematics. (2003). Principles and standards for school mathematics. Reston, VA: Author.

National Science Teachers Association. (2014). Science and engineering practice. Retrieved from http://ngss.nsta.org/PracticesFull.aspx

Rubenstein, R. N. (2013). Focused strategies for middle-grades mathematics vocabulary development. Mathematics Teaching in the Middle School, 13(4), 200-207.

Steele, D.F. (1999). Research into practice: Learning mathematical language in the zone of proximal development. Teaching Children Mathematics, 6(1), 38-42.

Thomas, J. W. (2000). A review of research on project-based learning. Retrieved from http://www.newtechnetwork.org.590elmp01.blackmesh.com/sites/default/files/dr/pblresearch2.pdf

Thompson, D., & Rubenstein, R. N. (2000). Learning mathematics vocabulary: Potential pitfallsand instructional strategies. Mathematics Teacher, 93(7), 568-574.

Vacca, R. T., & Vacca, A. L. (1996). Content area reading (5th ed.). New York, NY: Harper Collins/College. Yore, L.D., Craig, T. C., & Maguire, T.O. (1998). Index of science reading awareness: An interactive-constructive model, test verification, and grades 4-8 results. Journal of Research in Science Teaching, 35(1), 27-51.


International Journal of Contemporary Educational Research (IJCER)

www.ijcer.net

# Developing a Macroscopic Lens Into Middle School Reform: Psychometric Properties of the AMLE SIA

**Ayse Tugba Oner[1], Bilgin Navruz[1], and Robert M. Capraro[1]**
[1]Texas A&M University

**To cite this article:**

Oner, A.T., Navruz, B., & Capraro, R.M. (2015). Developing a macroscopic lens ınto middle school reform: psychometric properties of the AMLE SIA. *International Journal of Contemporary Educational Research, 2*(2), 89-103.

IJCER International Journal of Contemporary Educational Research
Volume 2, Number 2, July 2015, Page 89-103 ISSN: 2148-3868

# Developing a Macroscopic Lens Into Middle School Reform: Psychometric Properties of the AMLE SIA

**Ayse Tugba Oner[1*], Bilgin Navruz, Robert M. Capraro[1]**
[1]Texas A&M University

## Abstract

The purpose of this study was to assess the validity and reliability of the AMLE SIA, which was developed by the NMSA to provide the best educational programs for young adolescents to improve their skills. To promote these skills, NMSA suggested that schools needs to implement 16 characteristics nested within three categories. However, many middle schools failed to implement the practices. The reason might be the instrument itself due to including 96 items and the design. Therefore, the validity of the instrument was analyzed; response organization system was redesigned; the items were revised and eliminated by using regression (83 items); and the final instrument's validity was analyzed by using EFA (73% of variance explained) and the reliability (.98) was calculated.

**Key words:** School Improvement Assessment, Validity, Reliability

## Introduction

Young adults' future life is likely to be affected by decisions made between the ages of 10 and 15. Students decide about next steps during middle-level school years and school environment is an important component in their lifetime decision. Therefore, educational programs for young adolescents need to represent the best environment for 10-to 15- year- olds (National Middle School Association, 2010). In middle school, best practices have a positive effect on student achievement (Cook, Faulkner, & Kinne, 2009; Jackson & Lunenburg, 2010; McEwin & Greene, 2011; Mertens & Flowers, 2006). However, student achievement should not be the only concern; educators need to be aware of other aspects. Middle level education occurs during crucial years in students' educational lives because during these "transitional years" (National Middle School Association, 2010) students are likely to undergo significant changes in many aspects: physical, intellectual, moral, psychological, and socio-emotional. Any unwanted situation could affect these aspects, which could result in poor high school performance and high school dropout (National Middle School Association, 2010), which educators would not like to encounter. Thus, young adults needed an education that would improve their aspects and lead them to be optimistic about the future (National Middle School Association, 2014).

Students need an education that prepares them to overcome the present century's difficulties. Success in the ever-changing world is one factor that needs to be addressed in good education (National Middle School Association, 2014). The changing world causes changes in education, and educators must continue developing and trying to sustain the success of middle grade schools. During this development, educators use the attributes and characteristics recommended in *This We Believe: Keys to Educating Young Adolescents* (National Middle School Association, 2010), in which the National Middle School Association (NMSA) asserted knowledge and many skills that students should have to be self-actualized, fully functioning people. These skills and knowledge were major goals of middle level educators. For instance, students need to think critically and rationally; be able to gather, assess, and interpret data deeply; develop skills and interests; use digital tools to gather information from different sources; take responsibilities and make ethical decisions about their own health and wellness. To achieve these goals, educational programs should be developmentally responsive, challenging, empowering, and equitable (National Middle School Association, 2010), so students can overcome difficulties.

To achieve these goals, educational programs need to be aligned with 16 characteristics that were asserted by NMSA. The 16 characteristics were grouped into three categories (National Middle School Association, 2014):

* Corresponding Author: *Ayse Tugba Oner, aysetugbaoner@tamu.edu*

1) Curriculum, Instruction, and Assessment; 2) Leadership and Organization; and 3) Culture and Community. The Curriculum, Instruction, and Assessment Category dealt with having an engaged, active, purposeful, challenging, integrated learning and teaching environment and assessment. The Leadership and Organization Category concerned supporting professional development, collaboration, and organization for all stakeholders. The Culture and Community Category dealt with giving support and guidance to students by having a safe and inviting school environment. Having practices aligned with these three categories was useful for schools in terms of reintroducing successful and promising young adults to society. For instance, research conducted in a middle school in Kentucky showed that if a middle school concept aligned with the guidelines provided by *This We Believe: Keys to Educating Young Adolescents*, the students' academic achievement was higher than that of students in schools not aligned with the guidelines (Cook et al., 2009).

To assess whether a school's program aligns with these categories, NMSA introduced an instrument, the Association for Middle Level Education (AMLE) School Improvement Assessment (SIA), for middle school educators to evaluate middle schools. The AMLE SIA was based on 16 characteristics of *This We Believe: Keys to Educating Young Adolescents.* The SIA also provided reports about strengths and weaknesses of schools. By knowing the strengths and weaknesses of schools, teachers had a chance to help young adolescents become successful and responsible global citizens that were aimed in *This We Believe: Keys to Educating Young Adolescents.*

Even if *This We Believe: Keys to Educating Young Adolescents* presented the best practices in middle-level education, many middle level schools failed to fully implement middle-level practices (McEwin & Greene, 2011). One possible reason for this is that the instrument prepared by NMSA might not provide sufficient information about schools and their needs, which could be the result of an assessment that does not measure what it intend to assess. Therefore, the investigation of the SIA could be helpful in providing a valid instrument scores for evaluating both best practices in middle-level education and schools in terms of categories provided by NMSA (2011).

Assessing the assessment itself could be the first step before using that assessment. There are important questions that need to be answered in research, such as whether the assessment's content measures the intended purpose of the assessment and includes all related contents within it, how well the assessment predicts the criterion, and whether the measure assess the construct that is intended to represent. These questions deal with the validity of an assessment. Validity focuses on "whether a particular inference or conclusion is correct, reasonable, or accurate" (Bryant, 2000, p.101). Nunnally (1967) explicitly defined construct validity - one type of validity - as

> The degree to which it is necessary and difficult to validate measures of psychological variables is proportional to the degree to which the variable is concrete and abstract….To the extent that a variable is abstract rather than concrete, we speak of it as being a construct. Such a variable is…something that does not exist as an isolated, observable dimension of behavior. (pp. 84-85)

Therefore, it is important to have a measurement that can be accountable in terms of validity of constructs.

The purpose of this research was to analyze the validity and reliability of the Association for Middle Level Education (AMLE) School Improvement Assessment (SIA) and improve the organization of the SIA for future application. The reason was that there were concerns about the instrument, such as items in the assessment did not align well with the 16 Characteristics, or it was wordy and subject to spurious interpretation. Also the AMLE SIA instrument had 96 items; evaluating 96 items would be time consuming. As such, we also analyzed how effective the response organization system was and what would be the most efficient way to answer items in this instrument. By re-evaluating the instrument, we expected to have an instrument better aligned with the 16 characteristics, more trustworthy and valuable information about middle-grade practices, and a more focused instrument that would enable more targeted professional development and focused school improvement efforts.

## Method

The study consisted of four phases. The two of the phases were designed to provide insights into the validity of the scores obtained from the instrument, and other two of the phases focused on the instrument's functioning and sample results.

**Phase 1**

During Phase 1, several forms of validity were considered. Initially, Face Validity (Mosier, 1947) was used. With face validity, we checked to see if the operationalization, on its face, was a satisfactory translation of the construct we intended to measure. The intended construct was the "*This We Believe*. . ." document (National Middle School Association, 2010), which describes three categories comprised of 16 characteristics across 96 items. The 96 items were provided to 25 experts who examined the items and compared them to the statements in *This We Believe*. However, face validity is likely the weakest way to demonstrate construct validity (cf. Cronbach, & Meehl, 1955) and probably the most abused. Therefore, we took a very systematic and clinical approach. Experts were selected from the field of middle school were well versed in the constructs expressed in *This We Believe*. All the experts either were original authors of the document, reviewers of the document, or on the research advisory board for the National Middle School Associations, (has since changed its name to Association of Middle Level Educators [AMLE]).

*Face validity*. Face validity concerns the face of the instrument. If the instrument seems like a good measurement for the intended purpose, then it would have face validity. For instance, if an instrument is designed to measure math ability, as a researcher you might look at items and think that these items fit the instruments purpose. However, even if items are related to measuring math ability, because this is a subjective decision this type of validity might be considered as weak. The weakness of evidence does not mean that the instrument does not measure correctly; on the contrary, it might, but it still would be subjective. To prevent subjectivity, the researcher could send the instrument to other experts to improve its face validity. Therefore, besides face validity there are other validity types that should be examined to ensure the validity of an instrument.

*Content validity*. Content validity deals with whether an instrument evaluates relevant aspects of it (Bryant, 2000). The important issue for content validity is itself, because the test might measure a content related with the intended content which makes the test fail to have the content validity. The test also should measure all components of the content. For instance, a final exam for a course needs to measure the concepts that were taught in that course. Content validity is crucial because for some constructs, such as our final exam example, it might be easy to determine the criteria that fit to that construct, but for other constructs (e.g., attitude, intelligence) it is not (Trochim, 2006). The criterion in content validity is the measure itself. Therefore, there cannot be a correlation between a new measure and the criterion (Nunnally, 1967).

*Predictive validity*. The purpose of predictive validity is to "estimate some important form of behavior" (Nunnally, 1967), which is criterion. A test needs to be able to assess how successful it is in predicting the criterion before measuring it (Bryant, 2000). In that case, the instrument, theoretically, needs to predict the criterion. If there is a high correlation between predictor and criterion, the measure can predict something well. For instance, if the predictive validity of a measurement used to predict how well students perform in a specific grade is good, it should be able to predict students' success in that specific grade, not something else.

*Concurrent Validity*. In concurrent validity, unlike predictive validity, the scores are obtained from a new measure and criterion measure at the same time (Bryant, 2000). As in predictive validity, in concurrent validity, there should be a high correlation between these two measures. Thus, one test can be substituted for the other one because one measurement would be expected to measure the same construct. For instance, if there is a strong correlation between students' course grade and the passing test, the course-passing test could be given to students instead of the course itself. To have strong concurrent validity for an instrument, it would be better to collect test scores and criterion measure separately (Bryant, 2000).

**Phase 2**

*Response organization redesign*. It is important to pay as much attention to the online structure of the instrument as to the psychometric structure of the instrument. The instrument was originally designed to be administered one item at a time nested within characteristic, with characteristics nested within categories. This model required participants to be aware and to be presented with repetitious text about the category and characteristic to which they were responding. To examine the online structure of the instrument we considered the time it took to complete the instrument, time spent per item, and ways to improve the response time. Initially, we plotted the average item response time. For this part, there were 24 middle grades teachers, staff, and administrators. Then, we randomized the items to determine if there was any pattern with regard to the order of the characteristics and categories. A new sample from a different middle school was selected (n= 22). The reason for the randomization of items was to rule out whether the average response time was a factor of the item's difficulty toward the end of the test or due to the length. Finally, we revised the online layout and item response format again examining average item response time. Items were again nested within characteristics and characteristics nested within categories; however, now all of the items for a characteristic were displayed at the same time within the same block, and sliders were used for input. This effectively reduced the number of screens to 16 and eliminated the need to scroll. For this portion, a new sample of middle grades teachers, staff, and administrators from another middle school were selected (n= 23). The total number of items examined was 96.

**Phase 3**

This phase consists of two parts, a 15-member panel of experts and a separate group of six experts to review the comments and ratings for the purpose of deleting or revising items. In this phase, to have a more robust instrument, we analyzed it according to responses and comments from 15 middle grades experts. All experts were either classroom teachers (6) with 15 or more years of teaching and participation in AMLE or university middle level teacher educators (9) with 20 or more years combined of experience teaching middle school and university experiences researching middle level. They rated the items for relevance to the category and to the characteristic. Participants rated the items on a 100-point relevance rating scale where 1 indicated that the item did not align with the characteristic and category and 100 indicated that the item aligned well. They also were asked to provide qualitative commentary for each item. The qualitative commentary they were asked to cover concerned whether the item was relevant, possible rewording options, redundancy, or should be removed. For each item rated 50 or below, a new box would appear asking for a revision of the item or if the item should be removed from the scale and a reason.

Primarily, in the original AMLE SIA, there were 96 items. The first Category, Curriculum, Instruction, and Assessment, included 5 characteristics with 39 items. The Leadership and Organization Category consisted of 5 characteristics with 27 items. The last Category, Culture and Community, was comprised of 6 characteristics with 30 items.

All scores were converted to a percentage scale to make interpretation of data easier. Data were analyzed by aggregating scores across raters and computing means and standard deviations. Regression was used, and standardized $\beta$ weights were obtained with the total characteristic score as dependent and weighted item ratings as predictors. Structure coefficients were computed by correlating predicted values with the dependent variable. The structure coefficient estimates the percentage of useful variance accounted for by each item independent of other items. We used this as a relative measure of the importance of each item without considering whether or not that variance accounted for estimate was unique.

A separate and distinct group of six experts met to make final decisions for item revision or deletion. After consideration of quantitative and qualitative information, items with a rating of 90 or above were decided as adequate and retained; items with a rating between 70 and 90 were considered strong candidates for revision; and items with less than 70 were considered strong candidates for removal. In addition, if two items had equivalent ratings and the qualitative information indicated that the two items were addressing the same information, the item that had the higher $\beta$ weight was considered for revision whereas the other item was deleted. As a result, for the first Category, seven items were removed, eight items were selected for revision, and 24 items performed well. Three items from the second Category and three items from the third were eliminated. In the second Category, five and 19 items were considered for revision and retention, respectively. Seven items were selected for revision and 20 items performed well for the last Category.

**Phase 4**

*EFA analysis.* The sample included 15 teachers, nine of which were female. Seven of 15 individuals completed the survey by answering all the items.The final form of AMLE SIA included 83 items and was administered on the Web. Participants responded to items using a 0-to-100 unnumbered graphic rating scale. A 0 indicated never and 100 indicated all the time. Users moved a slider to indicate their response in the Qualtrics online survey software. The answer could be any number on the 0-100 scale, such as 47, 78, or 99. We chose the unnumbered graphic rating scale because it has been shown to be a favorable method for collecting psychometrically reliable scores (Cook, Heath, Thompson, & Thompson, 2001) (see Figure 4).

In the data set ~14% was missing observations, so as the first step for analysis was to impute data using the expectation-maximization (EM) algorithm, which is one of the popular methods for dealing with missing data. The Proc MI procedure in SAS 9.3 was used to impute the data. The Proc MI procedure reads raw data with missing observations and provides maximum likelihood variance-covariance estimates with the vector of means.

The EM covariance matrix is an excellent approach to deal with missing data prior to exploratory factor analysis (EFA) applications (Graham, 2012).

EFA was conducted to examine construct validity of scores. In EFA studies, the desired sample size was suggested as more than 200 (e.g., Cattell, 1978; Guilford, 1954). However, there were simulation studies that showed that EFA also provides sufficient results with a small sample size between 10 and 50 (Mundfrom, Shaw, & Ke, 2005; Preacher & MacCallum, 2002). Recently, it was found that EFA yielded reliable results even when sample size was smaller than 50 (i.e. even smaller than 10 in well conditioned data) (de Winter, Dodou, & Wieringa, 2009). Therefore, even in this study, in which there were 15 cases, conducting EFA would not yield misleading results.

After we obtained our variance-covariance matrix, we conducted most of the parametric statistical methods by using the information provided in that matrix, rather than using raw data (Zientek & Thomson, 2009). When we use sufficient decimal places in the variance covariance matrix, or in the correlation matrix with standard deviations of those variables, we would exactly estimate the same results, which can be found by reading raw data. Thus, we can use EM estimated variance covariance matrix in our EFA analysis to estimate our parameters of interest because EFA is one of the parametric multivariate models under the general linear model (GLM).

Principal components analysis was used as our factor extraction method. There were other methods to extract factors, such as principal axis factoring or maximum likelihood. Robust studies with smaller samples never used maximum likelihood as their method of choice. In terms of comparison of principal component and principal axis extraction methods, the two methods yield equivalent results when variables are increasing in the study or variables are reliable (Authors, 2004; Thompson, 1992). There were 83 items in the instrument, so we believe that either of these two choices would have returned the same results.

In EFA studies, the next decision is to determine the number of factors to extract from the correlation matrix. There are several strategies to decide how many factors should be extracted, such as Kaiser's eigenvalue greater than one rule, Catter's scree plot, and parallel analysis (Authors, 2004). The most commonly used method is the eigenvalue greater than one rule because this rule is default in many statistical packages (e.g., SPSS) (Thompson & Daniel, 1997). In our EFAs, we specified the number of factors as three because the instrument was based on middle grades educational theory.

Generally speaking, in EFA, after extracting the factors, rotation is almost always necessary to interpret factors easily (Thompson, 2004). There are two extraction methods: orthogonal and oblique. The difference between these rotation methods is that, philosophically, if the researcher believes the factors to be uncorrelated the orthogonal rotation retains uncorrelated characteristic in factors. For the oblique rotation the researcher must believe that the factors are correlated and the rotation retains that characteristic. Both rotation strategies are designed to obtain simple structure (Thurstone, 1947) for easier interpretation.

Correlated factors are one indicator for the existence of higher order factors (Gorsuch, 1983). While first order factors are extracted from the variable correlation matrix, higher order factors are extracted from the inter-factor correlation matrix (Thomson, 2004). Gorsuch (1983) explained higher order factors:

> Rotating obliquely in factor analysis implies that the factors do overlap and that there are, therefore, broader areas of generalizability than just a primary factor. Implicit in all oblique rotations are higher-order factors. It is recommended that these be extracted and examined so that the investigator may gain the fullest possible understanding of the data. (p. 255)

This portion of the study included two parts. In the first part, the authors reviewed the items, and they expected three correlated factors would be extracted under one higher order factor. We extracted three first-order factors, and one higher order factor was extracted from the "three by three by three" inter factor correlation matrix. Promax rotation, one of the oblique rotation strategies, was used to rotate three first order factors for easier interpretation. In the second part, we conducted three separate EFAs for each of the three categories separately. In each EFA, we extracted one factor from the inter-variable correlation matrix by using principal component analysis.

## Results

The results section moves from the evidence of score validity, examination of the response method to item revision, and finally exploratory factor analysis of the revised instrument.

**Results of Phase 1: Validity**

The expert review of face validity was in agreement that the items reflected the constructs expressed in the *This We Believe* . . . document. The rating was 100% agreement for the items in Characteristics 1, 6, 10, 11, 13, 15, 16, and 92% agreement for the items in Characteristics 2, 3, 4, 5, 6, 7, 8, 9, 12, 13, 14. However, there was more variability in Characteristics within categories. For characteristics within category there was 88% agreement for Category 1, 91% agreement for Category 2, and 84% for Category 3. The major conflict in agreement when considering the higher-level construct was that some characteristics could belong to two of the categories depending on the interpretations of the people answering the items. Primarily, the expert conversations were dominated by concerns for individual items comprising characteristics that could lead to the erroneous interpretations. Therefore, the problematic items were either removed or revised depending on consensus. After alterations and deletions, the group had 100% agreement across all items, characteristics, and categories.

For the concurrent validity we assessed the ability of the instrument to distinguish between those who knew the principles of *This We Believe* . . . from those who were unfamiliar. We compared the responses from the expert panel to those of 16 high school teachers who had never been middle school teachers, were not members of the Association of Middle Level Educators, and had not read the *This We Believe* . . . document. The two groups responded to the items and chose a 1 or 0, 1 indicating it was contained in the *This We Believe* . . . document and a 0 indicating that it was not. The score discrepancy was stark. Scores for the expert panel were very high (mean = 4.71, SD= .43) while the high school teachers were lower (mean = 3.15, SD= .99). Some potential reasons for why the score discrepancy was not even greater could lie in the fact that some of the principles are closely tied to good educational principles that cannot be isolated to the middle school only. However, when the groups were asked to determine if the items were contained in the *This We Believe* . . . document the difference was more dramatic. The chance score for the rating was .50 (there were only two choices, either a 0 or a 1, so participants had a 50/50 chance of guessing it correctly. The par score was .92, or 92 percent of the items were part of the document. The high school teachers' rating was .54, just above chance but well below the par score. The experts' score was .88. While they were able to identify all the distractors with 100% accuracy, they misclassified some items intended to measure the *This We Believe* . . . document. The identified items were the items that were flagged earlier as potentially problematic and subsequently either dropped or revised. The magnitude of the effect was .34, a sizeable difference between the groups.

**Results of Phase 2: Response Organization Redesign**

According to the results from 24 participants for items grouped within characteristic and characteristics within categories, participants' mean administration time was 34 minutes and 20 seconds. Participants' item response time grew longer toward the end of the assessment (see Figure 1). The assessment's items were not longer toward the end nor were they more complex; rather, respondents might suffer fatigue because of the total number of items (96) or the presentation of the items. As such, their response time was unacceptably long (see Figure 1) and calls into question the dependability of their responses toward the end of the administration.

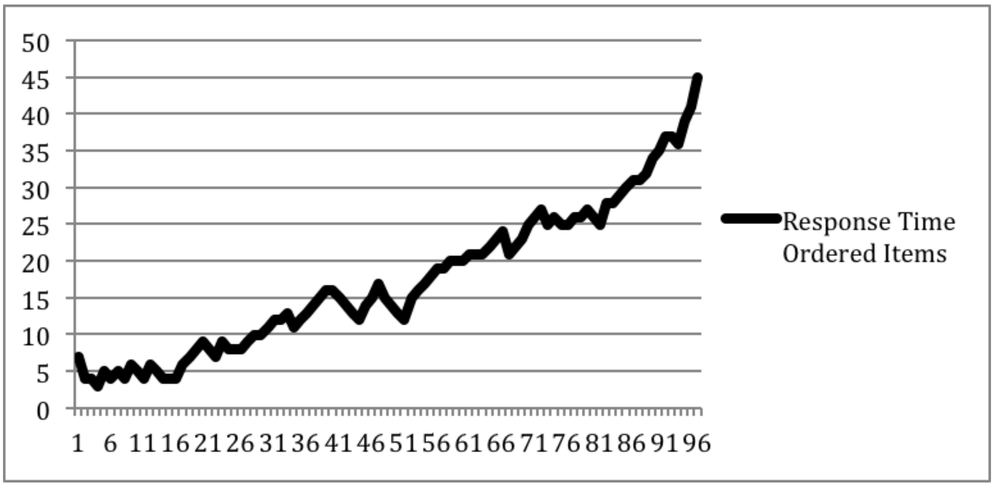



*Figure 1*. Response time ordered items

The results for the second type of response organization, random ordered items (not grouped by category or characteristic), showed that it was harder to keep the same response time performance when items were organized randomly. To complete the item the respondent was required to reach the descriptor for the category and the characteristic for every item. The mean administration time was 35 minutes and 44 seconds. The response time of participants was longer than the first organization type (see Figure 2). Organizing items randomly might create a chaotic environment for respondents. When two figures of two types of organization were taken into consideration (Figure 1 and Figure 2), the random ordered organization system graphic appeared more jagged in comparison to ordered items' graphic.

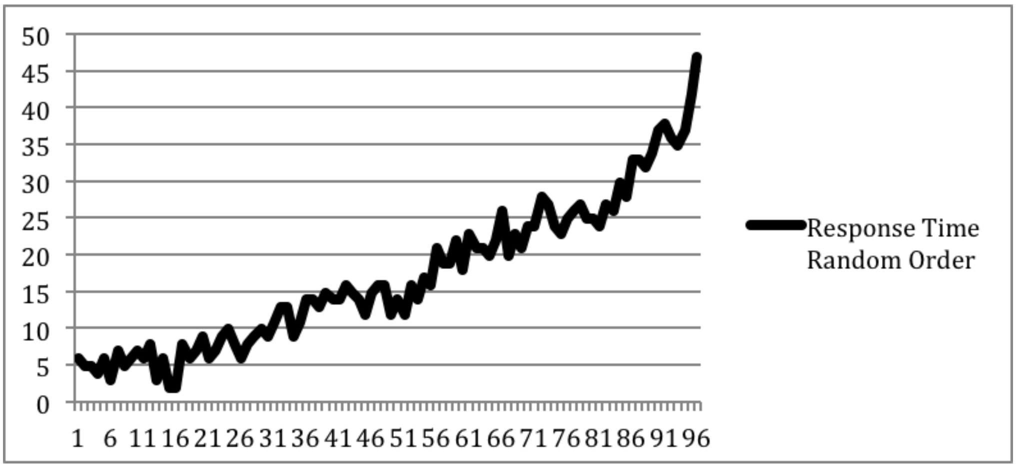



*Figure 2*. Response time random ordered items by standard presentation

The result for the systematic slider organization, items grouped by characteristic within category by slider, was the best one for respondents' (see Figure 3). The minimum and maximum response times per item were 3 and 13 minutes, respectively. The average administration time was 8 minutes.

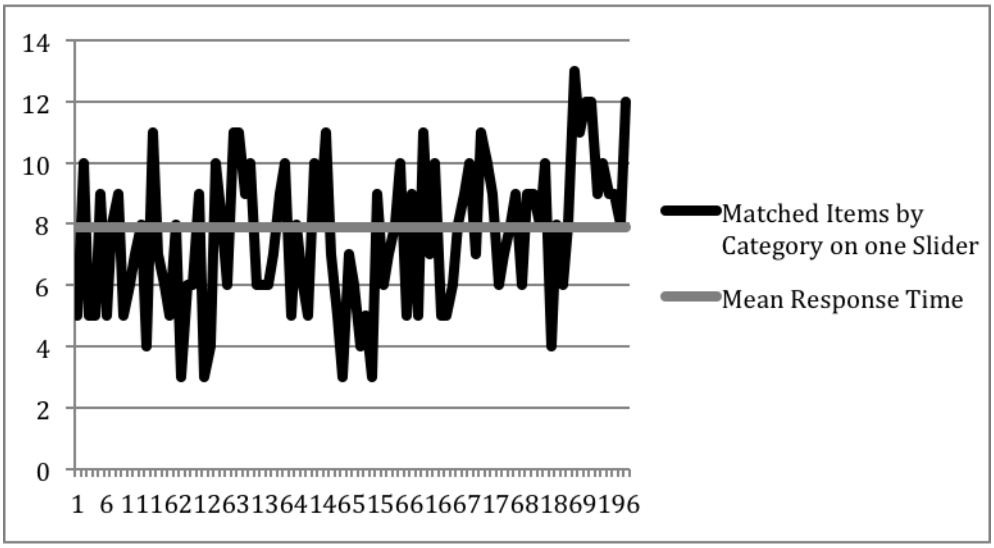



*Figure 3*. Response time ordered items grouped by slider choice

In this organization, the direction appeared only once per characteristic (see Figure 4). Also, this organization moved to the idea of 100-point scale, which most school personnel were very accustomed. The progress bar at the bottom of the electronic response page divided into 16 segments as opposed to 96. According to these results, it is important to organize items into one category, and present these items together in an online page in order to help respondents. Formatting the assessment plays a crucial role for the target group. It is important to notice that the variation around the mean was consistent and there was not quadratic appearance to the response time as the test time progressed. In addition, the average completion time was decreased to one-fourth of the time necessary.

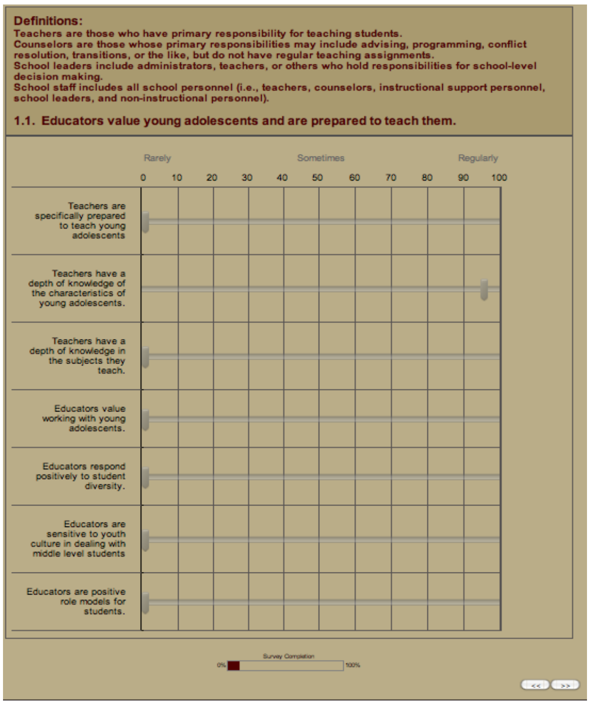


*Figure 4*. Example of items grouped by slider choice

**Results of Phase 4: EFA Analyses**

*One-first-order factor solution*

*Score reliability.* The reliability of scores was computed by Cronbach's α internal consistency coefficient. The reliability coefficient of whole test scores was 0.983. The internal consistency coefficient statistically showed the proportion of true score variance, so closer to 1 is desired. Our estimated reliability coefficient, 0.983, was very close to 1, thus we could say that scores were reliable.
*Factor structure.* To evaluate the test score validity, principal component analysis was conducted, following the guidelines described in Thompson and Daniel (1997).

*Three first order factor solution*

*Factor structure***.** The three eigenvalues prior to rotation for the factors subsequently were 44.861, 8.098, and 7.521, respectively. These three factors explained around 73% of the total variance of the variable correlation matrix. Because the authors expected the existence of one second-order factor, promax rotation for the simple structure was applied. Second order factors can only be extracted from an inter-factor correlation matrix after one of the oblique rotation strategies is used (Gorsuch, 1983; Thomspon, 2004). The eigenvalues for the components were, respectively, 26.322, 22.382, and 11.775 after rotation, and as expected they explained exactly the same amount of variance, 73%, of the variable correlation matrix. Table 1 shows pattern coefficients after Promax rotation was applied.

When one of the oblique rotations was applied to obtain simple structure, structure coefficients were not equal to pattern coefficients anymore, which is the case in orthogonal rotation (Thompson, 2004). Thus, whenever an oblique rotation is applied, structure coefficients should be reported separately (Thompson, 2004). Table 2 shows the structure coefficients.

Table 1. *Promax Rotated Pattern Coefficients*

| Item | Factor 1 | Factor 2 | Factor 3 | Item | Factor 1 | Factor 2 | Factor 3 | Item | Factor 1 | Factor 2 | Factor 3 | Item | Factor 1 | Factor 2 | Factor 3 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | **0.620** | 0.100 | 0.259 | 22 | **0.513** | 0.543 | -0.111 | 43 | **0.480** | 0.398 | 0.110 | 64 | 0.477 | -0.025 | **0.613** |
| 2 | -0.032 | **0.454** | 0.029 | 23 | **0.508** | 0.479 | 0.013 | 44 | 0.327 | 0.280 | **0.548** | 65 | **0.688** | 0.366 | -0.039 |
| 3 | 0.160 | 0.369 | -0.057 | 24 | 0.152 | **0.833** | -0.025 | 45 | -0.037 | **0.421** | 0.709 | 66 | -0.369 | 0.341 | **0.775** |
| 4 | 0.342 | -0.103 | **-0.643** | 25 | **0.455** | 0.579 | 0.038 | 46 | 0.336 | -0.257 | **0.586** | 67 | 0.106 | 0.008 | **0.817** |
| 5 | 0.193 | -0.040 | 0.085 | 26 | 0.073 | **0.749** | 0.313 | 47 | 0.212 | **0.830** | -0.046 | 68 | -0.049 | -0.338 | **0.916** |
| 6 | 0.375 | **0.467** | -0.262 | 27 | **0.941** | 0.063 | -0.483 | 48 | 0.259 | **0.817** | -0.141 | 69 | -0.651 | 0.245 | **0.660** |
| 7 | -0.237 | **0.466** | -0.391 | 28 | **0.734** | 0.375 | -0.121 | 49 | -0.037 | **0.974** | 0.023 | 70 | **0.564** | 0.242 | -0.011 |
| 8 | 0.344 | **0.480** | -0.187 | 29 | **0.886** | 0.128 | -0.034 | 50 | **0.625** | 0.390 | 0.023 | 71 | **0.682** | 0.155 | 0.027 |
| 9 | **0.688** | 0.131 | -0.537 | 30 | **1.042** | -0.044 | -0.099 | 51 | 0.088 | **0.524** | -0.673 | 72 | -0.098 | 0.469 | **0.545** |
| 10 | **1.170** | -0.406 | -0.077 | 31 | **0.962** | -0.111 | -0.025 | 52 | -0.065 | 0.301 | **0.617** | 73 | 0.212 | -0.071 | **0.862** |
| 11 | **0.537** | 0.117 | -0.607 | 32 | **0.956** | 0.095 | -0.195 | 53 | 0.389 | -0.155 | **0.515** | 74 | **0.583** | 0.301 | 0.305 |
| 12 | **0.960** | 0.024 | -0.204 | 33 | **0.799** | 0.072 | 0.167 | 54 | 0.022 | **0.852** | 0.174 | 75 | **0.740** | 0.149 | 0.254 |
| 13 | **0.961** | -0.082 | 0.049 | 34 | 0.443 | **0.573** | 0.060 | 55 | 0.018 | **0.951** | -0.095 | 76 | 0.053 | **0.768** | 0.259 |
| 14 | **0.589** | 0.456 | -0.037 | 35 | 0.450 | **0.469** | 0.206 | 56 | -0.157 | **1.021** | -0.457 | 77 | **0.575** | 0.314 | 0.187 |
| 15 | **0.900** | 0.125 | -0.178 | 36 | 0.391 | **0.634** | 0.053 | 57 | -0.119 | **0.815** | 0.240 | 78 | 0.286 | 0.194 | **0.654** |
| 16 | **0.940** | -0.391 | 0.343 | 37 | 0.935 | **-0.618** | 0.408 | 58 | -0.245 | 0.398 | 0.336 | 79 | -0.029 | **0.903** | 0.179 |
| 17 | **0.467** | 0.301 | -0.350 | 38 | 0.294 | **0.692** | 0.057 | 59 | -0.307 | **0.903** | -0.067 | 80 | **0.661** | -0.147 | 0.276 |
| 18 | **0.502** | 0.583 | -0.069 | 39 | 0.400 | **0.548** | 0.239 | 60 | -0.352 | **0.900** | 0.210 | 81 | **0.492** | 0.156 | 0.273 |
| 19 | **1.001** | -0.255 | 0.008 | 40 | 0.304 | **0.432** | 0.453 | 61 | -0.175 | 0.379 | **0.638** | 82 | 0.498 | -0.047 | **0.548** |
| 20 | **0.977** | -0.212 | 0.003 | 41 | **0.792** | 0.114 | 0.062 | 62 | -0.148 | **1.069** | -0.014 | 83 | **0.474** | 0.273 | 0.343 |
| 21 | **0.724** | 0.241 | -0.244 | 42 | **0.602** | 0.091 | -0.182 | 63 | -0.163 | **1.119** | -0.125 | | | | |

*Note.* Factor pattern coefficients greater than |.4| are bolded.

Table 2. *Structure Coefficients*

| Item | Factor 1 | Factor 2 | Factor 3 | Item | Factor 1 | Factor 2 | Factor 3 | Item | Factor 1 | Factor 2 | Factor 3 | Item | Factor 1 | Factor 2 | Factor 3 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.794 | 0.587 | 0.565 | 22 | 0.804 | 0.820 | 0.321 | 43 | 0.776 | 0.740 | 0.471 | 64 | 0.725 | 0.509 | 0.808 |
| 2 | 0.263 | 0.445 | 0.192 | 23 | 0.812 | 0.801 | 0.418 | 44 | 0.737 | 0.696 | 0.798 | 65 | 0.898 | 0.779 | 0.398 |
| 3 | 0.365 | 0.446 | 0.155 | 24 | 0.661 | 0.919 | 0.364 | 45 | 0.530 | 0.673 | 0.856 | 66 | 0.176 | 0.412 | 0.748 |
| 4 | 0.001 | -0.140 | -0.536 | 25 | 0.832 | 0.877 | 0.458 | 46 | 0.428 | 0.180 | 0.631 | 67 | 0.463 | 0.391 | 0.866 |
| 5 | 0.205 | 0.114 | 0.153 | 26 | 0.675 | 0.917 | 0.636 | 47 | 0.709 | 0.944 | 0.367 | 68 | 0.135 | -0.013 | 0.764 |
| 6 | 0.553 | 0.599 | 0.080 | 27 | 0.773 | 0.462 | -0.053 | 48 | 0.707 | 0.923 | 0.287 | 69 | -0.215 | 0.096 | 0.475 |
| 7 | -0.115 | 0.167 | -0.312 | 28 | 0.915 | 0.785 | 0.340 | 49 | 0.579 | 0.960 | 0.385 | 70 | 0.710 | 0.589 | 0.325 |
| 8 | 0.562 | 0.621 | 0.147 | 29 | 0.951 | 0.666 | 0.397 | 50 | 0.877 | 0.788 | 0.443 | 71 | 0.789 | 0.590 | 0.380 |
| 9 | 0.538 | 0.350 | -0.191 | 30 | 0.973 | 0.567 | 0.333 | 51 | 0.125 | 0.318 | -0.431 | 72 | 0.429 | 0.620 | 0.685 |
| 10 | 0.884 | 0.293 | 0.269 | 31 | 0.881 | 0.478 | 0.345 | 52 | 0.388 | 0.500 | 0.706 | 73 | 0.538 | 0.396 | 0.926 |
| 11 | 0.349 | 0.216 | -0.331 | 32 | 0.932 | 0.615 | 0.253 | 53 | 0.514 | 0.288 | 0.622 | 74 | 0.902 | 0.783 | 0.673 |
| 12 | 0.887 | 0.543 | 0.218 | 33 | 0.915 | 0.634 | 0.538 | 54 | 0.627 | 0.934 | 0.514 | 75 | 0.942 | 0.709 | 0.630 |
| 13 | 0.931 | 0.536 | 0.431 | 34 | 0.825 | 0.872 | 0.473 | 55 | 0.569 | 0.925 | 0.282 | 76 | 0.643 | 0.902 | 0.580 |
| 14 | 0.857 | 0.809 | 0.394 | 35 | 0.831 | 0.829 | 0.582 | 56 | 0.282 | 0.746 | -0.128 | 77 | 0.851 | 0.745 | 0.557 |
| 15 | 0.901 | 0.616 | 0.258 | 36 | 0.809 | 0.898 | 0.468 | 57 | 0.491 | 0.833 | 0.505 | 78 | 0.688 | 0.626 | 0.853 |
| 16 | 0.844 | 0.327 | 0.595 | 37 | 0.726 | 0.123 | 0.571 | 58 | 0.148 | 0.376 | 0.385 | 79 | 0.610 | 0.954 | 0.517 |
| 17 | 0.504 | 0.456 | -0.032 | 38 | 0.750 | 0.897 | 0.453 | 59 | 0.226 | 0.686 | 0.151 | 80 | 0.689 | 0.372 | 0.503 |
| 18 | 0.835 | 0.869 | 0.373 | 39 | 0.843 | 0.889 | 0.623 | 60 | 0.299 | 0.762 | 0.408 | 81 | 0.706 | 0.568 | 0.545 |
| 19 | 0.846 | 0.372 | 0.340 | 40 | 0.768 | 0.797 | 0.752 | 61 | 0.335 | 0.518 | 0.710 | 82 | 0.704 | 0.476 | 0.744 |
| 20 | 0.846 | 0.398 | 0.341 | 41 | 0.890 | 0.631 | 0.446 | 62 | 0.512 | 0.972 | 0.338 | 83 | 0.792 | 0.701 | 0.653 |
| 21 | 0.769 | 0.597 | 0.161 | 42 | 0.580 | 0.395 | 0.112 | 63 | 0.479 | 0.968 | 0.239 | | | | |

The Promax rotation produced correlated first-order factors, so these overlapped factors implied the existence of a higher order factor that could be extracted from the inter-factor correlation matrix. The second-order factor was extracted from the inter-factor correlation matrix. Table 3 contains the correlation matrix for Promax rotated first-order factors and the second order factor. The eigenvalue for the second-order factor was 1.969, and it explained 66% of the variance in the inter-factor correlation matrix.

Table 3. *Inter-Factor Correlation Matrix and Second-Order Factor Structure Coefficients*

| | Correlation Matrix | | | Second-Order Coefficients |
|---|---|---|---|---|
| | First1 | First2 | First3 | |
| First1 | 1 | | | 0.861 |
| First2 | 0.623 | 1 | | 0.842 |
| First3 | 0.430 | 0.388 | 1 | 0.720 |

First-order factors were not observed variables; they were abstractions of observed variables. Because second-order factors were extracted from the inter-factor correlation matrix, second-order factors were "abstractions of abstractions" (Thomson, 2004, p. 74). We would not want to interpret second-order factors in terms of abstractions (e.g., first-order factors); instead, we should interpret second-order factors in terms of measured variables. To be able to interpret second-order factors in terms of measured variables, the Schimid and Leiman (1957) solution is provided in Table 4.

Table 4. *Schmid and Leiman Solution*

| Item # | Second | First1 | First2 | First3 |
|---|---|---|---|---|
| 1 | 0.804 | 0.160 | 0.029 | 0.125 |
| 2 | 0.376 | -0.008 | 0.132 | 0.014 |
| 3 | 0.407 | 0.041 | 0.107 | -0.027 |
| 4 | -0.255 | 0.088 | -0.030 | -0.310 |
| 5 | 0.194 | 0.050 | -0.012 | 0.041 |
| 6 | 0.528 | 0.097 | 0.136 | -0.126 |
| 7 | -0.093 | -0.061 | 0.136 | -0.188 |
| 8 | 0.566 | 0.089 | 0.140 | -0.090 |
| 9 | 0.316 | 0.178 | 0.038 | -0.259 |
| 10 | 0.610 | 0.303 | -0.118 | -0.037 |
| 11 | 0.124 | 0.139 | 0.034 | -0.292 |
| 12 | 0.700 | 0.248 | 0.007 | -0.098 |
| 13 | 0.794 | 0.249 | -0.024 | 0.024 |
| 14 | 0.865 | 0.152 | 0.133 | -0.018 |
| 15 | 0.752 | 0.233 | 0.036 | -0.086 |
| 16 | 0.727 | 0.243 | -0.114 | 0.165 |
| 17 | 0.404 | 0.121 | 0.088 | -0.169 |
| 18 | 0.874 | 0.130 | 0.170 | -0.033 |
| 19 | 0.653 | 0.259 | -0.074 | 0.004 |
| 20 | 0.665 | 0.253 | -0.062 | 0.001 |
| 21 | 0.651 | 0.187 | 0.070 | -0.118 |
| 22 | 0.819 | 0.133 | 0.158 | -0.053 |
| 23 | 0.850 | 0.131 | 0.139 | 0.006 |
| 24 | 0.814 | 0.039 | 0.242 | -0.012 |
| 25 | 0.907 | 0.118 | 0.168 | 0.018 |
| 26 | 0.919 | 0.019 | 0.218 | 0.151 |
| 27 | 0.516 | 0.243 | 0.018 | -0.233 |
| 28 | 0.861 | 0.190 | 0.109 | -0.058 |
| 29 | 0.846 | 0.229 | 0.037 | -0.016 |
| 30 | 0.789 | 0.269 | -0.013 | -0.048 |
| 31 | 0.717 | 0.249 | -0.032 | -0.012 |
| 32 | 0.763 | 0.247 | 0.028 | -0.094 |
| 33 | 0.869 | 0.207 | 0.021 | 0.080 |
| 34 | 0.907 | 0.115 | 0.167 | 0.029 |
| 35 | 0.931 | 0.116 | 0.136 | 0.099 |
| 36 | 0.909 | 0.101 | 0.184 | 0.026 |
| 37 | 0.578 | 0.242 | -0.180 | 0.197 |
| 38 | 0.877 | 0.076 | 0.201 | 0.027 |
| 39 | 0.978 | 0.103 | 0.159 | 0.115 |
| 40 | 0.952 | 0.079 | 0.126 | 0.218 |
| 41 | 0.823 | 0.205 | 0.033 | 0.030 |
| 42 | 0.464 | 0.156 | 0.026 | -0.088 |
| 43 | 0.828 | 0.124 | 0.116 | 0.053 |
| 44 | 0.912 | 0.085 | 0.081 | 0.264 |
| 45 | 0.833 | -0.010 | 0.122 | 0.342 |
| 46 | 0.495 | 0.087 | -0.075 | 0.282 |
| 47 | 0.848 | 0.055 | 0.241 | -0.022 |
| 48 | 0.809 | 0.067 | 0.238 | -0.068 |
| 49 | 0.805 | -0.010 | 0.283 | 0.011 |
| 50 | 0.883 | 0.162 | 0.113 | 0.011 |
| 51 | 0.033 | 0.023 | 0.152 | -0.324 |
| 52 | 0.642 | -0.017 | 0.088 | 0.297 |
| 53 | 0.575 | 0.101 | -0.045 | 0.248 |
| 54 | 0.862 | 0.006 | 0.248 | 0.084 |
| 55 | 0.748 | 0.005 | 0.277 | -0.046 |
| 56 | 0.396 | -0.041 | 0.297 | -0.220 |
| 57 | 0.757 | -0.031 | 0.237 | 0.116 |
| 58 | 0.366 | -0.063 | 0.116 | 0.162 |
| 59 | 0.448 | -0.079 | 0.263 | -0.032 |
| 60 | 0.606 | -0.091 | 0.262 | 0.101 |
| 61 | 0.628 | -0.045 | 0.110 | 0.307 |
| 62 | 0.763 | -0.038 | 0.311 | -0.007 |
| 63 | 0.712 | -0.042 | 0.326 | -0.060 |
| 64 | 0.831 | 0.123 | -0.007 | 0.295 |
| 65 | 0.873 | 0.178 | 0.106 | -0.019 |
| 66 | 0.527 | -0.095 | 0.099 | 0.373 |
| 67 | 0.686 | 0.027 | 0.002 | 0.394 |
| 68 | 0.333 | -0.013 | -0.098 | 0.441 |
| 69 | 0.121 | -0.168 | 0.071 | 0.318 |
| 70 | 0.681 | 0.146 | 0.070 | -0.005 |
| 71 | 0.737 | 0.176 | 0.045 | 0.013 |
| 72 | 0.703 | -0.025 | 0.136 | 0.263 |
| 73 | 0.743 | 0.055 | -0.021 | 0.415 |
| 74 | 0.975 | 0.151 | 0.088 | 0.147 |
| 75 | 0.945 | 0.191 | 0.043 | 0.122 |
| 76 | 0.879 | 0.014 | 0.223 | 0.125 |
| 77 | 0.894 | 0.149 | 0.091 | 0.090 |
| 78 | 0.880 | 0.074 | 0.056 | 0.315 |
| 79 | 0.864 | -0.008 | 0.263 | 0.086 |
| 80 | 0.644 | 0.171 | -0.043 | 0.133 |
| 81 | 0.752 | 0.127 | 0.045 | 0.132 |
| 82 | 0.784 | 0.129 | -0.014 | 0.264 |
| 83 | 0.885 | 0.123 | 0.079 | 0.165 |

*Note.* Factor pattern coefficients lower than |.4| are underlined.

*Separate EFAs for each category*

*Reliability structure.* Internal consistency reliability estimates for each category were calculated. The reliability estimate for the first Category, which had 32 items, was 0.978. Cronbach's α for the second Category, which included 24 items, was 0.939, and the estimate for the third Category, which had 27 items, was 0.969. Although the third Category's reliability estimate was 0.939, it was very close to 1. The other two categories' estimates were close to 1; therefore, we could conclude that our scores were reliable for each category.

*Factor structure.* Table 5 shows the pattern/structure coefficients for each category. Three separate EFAs were conducted by using the principal components method. The first EFA for Category 1 with 32 items resulted in only 5 pattern/structure coefficients smaller than .4. A second separate EFA for Category 2 with 24 items (33-56) resulted in only 1 item (51) and had a value smaller than .4. The other separate EFA for Category 3 with 27 items produced only 1 item and had a value smaller than .4. All others had pattern/structure coefficients bigger than .4 absolute value.

Table 5. *Pattern/Structure Coefficient for Each Category*

| Category 1 | | Category 2 | | Category 3 | |
|---|---|---|---|---|---|
| **Item #** | Factor 1.1 | Item # | Factor 2.1 | Item # | Factor 3.1 |
| 1 | **0.792** | 33 | **0.869** | 57 | **0.651** |
| 2 | 0.357 | 34 | **0.963** | 58 | **0.558** |
| 3 | **0.432** | 35 | **0.985** | 59 | **0.534** |
| 4 | 0.013 | 36 | **0.975** | 60 | **0.594** |
| 5 | 0.273 | 37 | **0.489** | 61 | **0.719** |
| 6 | **0.636** | 38 | **0.908** | 62 | **0.741** |
| 7 | 0.025 | 39 | **0.963** | 63 | **0.704** |
| 8 | **0.683** | 40 | **0.956** | 64 | **0.841** |
| 9 | **0.611** | 41 | **0.875** | 65 | **0.815** |
| 10 | **0.817** | 42 | **0.546** | 66 | **0.626** |
| 11 | 0.353 | 43 | **0.886** | 67 | **0.735** |
| 12 | **0.846** | 44 | **0.898** | 68 | **0.411** |
| 13 | **0.929** | 45 | **0.797** | 69 | 0.146 |
| 14 | **0.923** | 46 | **0.467** | 70 | **0.791** |
| 15 | **0.914** | 47 | **0.900** | 71 | **0.816** |
| 16 | **0.758** | 48 | **0.884** | 72 | **0.815** |
| 17 | **0.524** | 49 | **0.844** | 73 | **0.810** |
| 18 | **0.904** | 50 | **0.864** | 74 | **0.933** |
| 19 | **0.768** | 51 | 0.155 | 75 | **0.901** |
| 20 | **0.824** | 52 | **0.671** | 76 | **0.895** |
| 21 | **0.829** | 53 | **0.598** | 77 | **0.910** |
| 22 | **0.902** | 54 | **0.897** | 78 | **0.895** |
| 23 | **0.852** | 55 | **0.825** | 79 | **0.793** |
| 24 | **0.772** | 56 | **0.511** | 80 | **0.709** |

(continued)

| **Item #** | Factor 1.1 | Item # | Factor 2.1 | Item # | Factor 3.1 |
|---|---|---|---|---|---|
| 25 | **0.889** | | | 81 | **0.833** |
| 26 | **0.759** | | | 82 | **0.870** |
| 27 | **0.800** | | | 83 | **0.946** |
| 28 | **0.950** | | | | |
| 29 | **0.963** | | | | |
| 30 | **0.931** | | | | |
| 31 | **0.888** | | | | |
| 32 | **0.915** | | | | |

*Note*. Factor pattern coefficients greater than |.4| are bolded.

The extracted factor from the 32 by 32 inter-variable correlation matrix for Category 1, explained around 58 % of variation. The other one factor extracted from 24 items for Category 1 explained 65 % of the variation. The factor extracted from 27 items inter-variable correlation matrix for Category 3, explained 58 % of the variation among variables.

## Discussion

The four phases of the study shed important light on the usefulness of the AMLE's SIA. The validity indicators are reasonable. The benchmarks by which the validly was judged are not the only ones available but do comprise prominent means for determining that the instrument does measure what it is intended to measure. Arguably, the strongest indicator for validity claims comes from the ratings of experts as compared to novices unfamiliar with the AMLE SIA or the principles on which it is based. The novice group was barely above chance score indicating that guessing could likely have accounted for the same score while the experts were much closer to par score. It is very important to note that the experts were able to pick the distractors out 100% of the time. The misclassification of items as not being representative of *This We Believe. . .* was a strong indicator for further analysis and instrument refinement.

The refinement of the instrument led to experts having to sit around a table, listen to each other explain their rationale for their ratings and to come to consensus on what to do with particular items. The analysis process reduced the number of items and increased the clarification of items that previously caused ambiguity.

Perhaps the most detrimental aspect to validity and reliability is instrument length. In the original version items were presented within characteristic and category but one item at a time. The process required 96 mouse clicks to progress through the instrument with additional clicks for consent, information, and terminology. When looking at the graphs there was not explanation for why as the items progressed it would take respondents more time to answer an item. So the items were randomized to determine if the items toward the end of the assessment were more time consuming. In fact, they were not and time to completion increased. We reduced the total number of mouse clicks to 16 with the same number of introductory clicks. Presenting a grouping of items by characteristic within the category reduced the amount of reading and allowed respondents to simply move a slider. They were able to see which items came before and after the item they were addressing. The perception could have been that there were fewer items. In fact, the time to completion of the instrument was reduced to ¼ the time needed previously and the reliability was very high at .98. Therefore, we can rule out that the respondents simply guessed without reading the prompts.

Finally, the EFA clearly indicates that the dropped items and the revised items provide a reasonable fit to the data, and the new instrument accounted for 73% of the variance, a very respectable number. The factor loadings were mostly as expected with a few items being problematic. Some items that we expected to load on Category 3 actually loaded on Category 1. The potential problem here is that respondents may have viewed the culture and community items as they relate to curriculum because the items do have implications for instruction.
The revised instrument does yield valid and reliable scores. However, more data is necessary for a fine-grained analysis. That data also needs to be more representative of middle school programs and the way the middle school concept is enacted in the U.S. While we believe the analysis is robust, the sample size for the EFA only

accounts for a single school district in a single state. The results seem to indicate that fewer items could improve the instrument, as could the collapsing of some of the characteristics.

## References

Bryant, F. (2000). Assessing the validity of measurement. In L. G. Grimm, & P. R. Yarnold (Eds.), *Reading and understanding more multivariate statistics* (pp. 99-139). Washington, D.C.: American Psychological Association.

Cattell, R. B. (1978). *The scientific use of factor analysis in behavioral and life sciences*. New York: Plenum.

Cook, C., Faulkner, S., & Kinne, L. (2009). Indicators of middle school implementation: How do Kentucky's schools to watch measure up? *Research in Middle Level Education Online*, *32*, 1-10.

Cook, C., Heath, F., Thompson, R. L., & Thompson, B. (2001). Score reliability in webor internet-based surveys: Unnumbered graphic rating scales versus likert-type scales. *Educational and Psychological Measurement*, *61*(4), 697-706.

de Winter, J. C. F., Dodou, D. & Wieringa, P. A. (2009). Exploratory factor analysis with small sample sizes. *Multivariate Behavioral Research, 44*(2), 147-181.

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual review of psychology*, *60*, 549-576.

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.

Guilford, J. P. (1954). Psychometric methods (2nd ed.). New York: McGraw-Hill.

Authors (2004).

Horn, J. L. (1965). A rationale and test for the number if factors in factor analysis. *Psychometrika*, *30*, 179-185.

Jackson, S., & Lunenburg, F. (2010). School performance indicators, accountability ratings, and student achievement. *American Secondary Education, 39*, 27-44.

McEwin, C. K., & Greene, M. W. (2011). *The status of programs and practices in America's middle schools: Results from two national studies*. Association for Middle Level Education. Retrieved from http://www.amle.org/portals/0/pdf/research/Research_from_the_Field/Status_Programs_Practices_AMLE.pdf.

Mertens, S. B., & Flowers, N. (2006). Middle Start's impact on comprehensive middle school reform. *Middle Grades Research Journal, 1*, 1-26.

Mundfrom, D. J., Shaw, D. G., & Ke, T. L. (2005). Minimum sample size recommendations for conducting factor analyses. *International Journal of Testing, 5*, 159-168.

National Middle School Association. (2010). *This we believe: Keys to educating young adolescents*. Westerville, OH: National Middle School Association.

National Middle School Association. (2014). *This we believe: Keys to educating young adolescents position paper of national middle school association executive summary*. Retrieved from http://www.uww.edu/Documents/colleges/coeps/academics/This_We_Believe_Exec_Summary.pdf

Nunnally, J. C. (1967). *Psychometric theory*. New York: McGraw-Hill.

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavioral Research Methods, Instrument, & Computers*, *32*(3), 396-402.

Preacher, K. J., & MacCallum, R. C. (2002). Exploratory factor analysis in behavior genetics research: Factor recovery with small sample sizes. *Behavior Genetics, 32*, 153-161.

Schimid, J., & Leiman, J. (1957). The development of hierarchical factor solutions. *Psychometrica*, *22*, 53-61.

Thompson, B. (1992). A partial test distribution for cosines among factors across samples, In B. Thompson (Ed.), *Advances in social science methodology* (Vol. 2, pp. 81-97). Greenwich, CT: JAI Press.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis*. Washington, DC: American Psychological Association.

Thompson, B., & Daniel, L. G. (1996). Factor analytic evidence for the construct validity of scores: A historical review and some guidelines. *Educational and Psychological Measurement, 56*(2), 197-208.

Thurstone, L. L. (1947). *Multiple factor analysis.* Chicago: University of Chicago Press.

Trochim, W. M. K. (2006). *Measurement validity types*. Research Methods Knowledge Base. Retrieved from http://www.socialresearchmethods.net/kb/measval.php

Zientek, L. R., & Thompson, B. (2009). Matrix summaries improve research reports: Secondary analyses using published literature. *Educational Researcher*, *38*(5), 343-352.


International Journal of Contemporary Educational Research (IJCER)

www.ijcer.net

# Effects of Reciprocal Teaching on EFL Fifth Graders' English Reading Ability

**Teng-lung Peng[1], Shu-hui Wang[1]**
[1] National Yunlin University of Science and Technology

**To cite this article:**

Peng, T. & Wang, S. (2015). Effects of reciprocal teaching on EFL fifth graders' English reading ability. *International Journal of Contemporary Educational Research*, *2*(2), 76-88

IJCER International Journal of Contemporary Educational Research
Volume 2, Number 2, July 2015, Page 76-88 ISSN: 2148-3868

# Effects of Reciprocal Teaching on EFL Fifth Graders' English Reading Ability

**Teng-lung Peng[1*], Shu-hui Wang[1]**
[1] National Yunlin University of Science and Technology

## Abstract

The purpose of this paper is to investigate the effects of Reciprocal Teaching on EFL fifth graders' reading ability, in terms of word recognition and reading comprehension in an elementary school in Taiwan. The participants in this research were fifty-three fifth-graders of an elementary school, 25 males and 28 females, in two intact classes in Taiwan. Students in one fifth-grade class, the control group, received regular English instruction, while those in the other fifth-grade class, the experimental group, received reciprocal teaching program, in two-hour English classes per week for 12 weeks, the duration of this study. Pretest-posttest of word recognition tests and reading comprehension tests were applied to evaluate students' progress of reading ability before and after reciprocal teaching program. Six students were interviewed after the program, and students' attitudes toward reciprocal teaching in English reading were recorded. The research results indicated that most students made prominent improvement in their English reading ability, word recognition, and reading comprehension. In addition, most participants had positive attitude toward reciprocal teaching, and they liked reciprocal teaching to be incorporated into English classes. Finally, based on the findings, some implications are also proposed to be of help to those who are English teachers or educational practitioners in elementary schools.

**Key words:** Reciprocal Teaching, EFL, Word recognition tests, Reading comprehension tests.

## Introduction

It is well acknowledged that students' reading ability can have important effects on their academic achievement. By and large, students with good reading ability perform better than those with less developed reading skills. A great deal of evidence shows that good readers are very strategic as they read. Strategically competent readers may well make predictions before they read, ask questions as they read, seek further clarification when they are confused, and write summaries of what they have read for themselves (Pressley, 1998). In summary, students' reading comprehension plays a vital role in their academic achievement.

However, students with limited skills in reading comprehension and low motivation for reading, pose challenges to instructors. These challenges may be intermingled when these students lack access to interesting materials and strategic activities and when they miss the concept of reading as a process of understanding meaning from texts (Fevre, Moore, & Wilkinson, 2003). In fact, these students are entitled to access to interesting materials and strategic activities which may get them actively engaged in reading even though the texts are cognitively demanding. Without this access, students with a reading deficiency may also result in a knowledge deficiency (Fielding & Roller, 1992; Jenkins, Stein, & Wysocki, 1984). Therefore, it is imperative for instructors to provide students with interesting materials and strategic activities incorporated into effective reading strategy instruction.

Existing research recognizes the critical role played by reciprocal teaching in improving students' reading comprehension. To date, there are few studies that have used reciprocal teaching to improve EFL elementary school students' ability to comprehend English text. Therefore, the purpose of this study is to investigate the effects of reciprocal teaching on Taiwanese EFL elementary school students' reading ability.

The study explores three principal research questions as follows:

* Corresponding Author: *Teng-lung Peng, antern.perng@gmail.com*

1. To what extent can students perform differently in word recognition before and after the implementation of reciprocal teaching instruction?
2. To what extent can students perform differently in reading comprehension before and after the implementation of reciprocal teaching instruction?
3. What are the effects of reciprocal teaching on enhancing fifth-grade students' attitudes toward English reading?

**Literature Review**

Since Palinscar and Brown (1984) introduced the concept of reciprocal teaching, a number of studies have been conducted to examine the influence of reciprocal teaching on students' ability to comprehend text. Three major findings emerged from their review of the literature (Rosenshine & Meister, 1993). The first finding was that a combination of explicit instruction and reciprocal teaching yielded more significant results than reciprocal teaching only. The second finding was that when all the students were taught, the results were usually significant. But when below-average students were taught, the results were mostly non-significant. The third finding was that when experimenter-developed tests were used, the results were usually significant. Yet, when standardized tests were used, the results were usually non-significant.

Traditional reciprocal teaching consisted of the following components: four steps of instruction (i.e., predicting, questioning, clarifying, and summarizing), a small-group classroom setting, students as leaders of discussion and teachers as facilitators. Traditionally, reciprocal teaching was carried out with small groups working independently, and students took on leading roles in reading lessons (Palinscar & Brown, 1984). Provision of guided practice was a major composition of reciprocal teaching. At the initial phase, the student was an observer and did little cognitive work. The teacher made use of the information provided by students as a form of informal assessment. Then, the teacher created instructional scaffolding based on this assessment (Alfassi, 1998; Duffy & Roehler, 1987). In brief, in traditional reciprocal teaching, students predict before reading, ask questions while reading, seek clarification when confused, and make a summary of what they have read after reading.

However, reciprocal teaching is flexible and may be modified, and instructors are not restricted to following a prescribed set of instructional routines (Oczkus, 2003). Reciprocal-teaching lessons presented were taught with the whole class as illustrated by Myers (2006). In addition, Palincsar and Brown (1984) suggested that heterogeneous groupings by age or by reading proficiency level may maximize the advantages of the reading process by offering students more effective peer models, apart from those models provided by the instructor. In view of classroom management, reciprocal teaching may be modified and carried out during read-aloud lessons.

Adaptations to a reciprocal teaching strategy may prove effective in improving young learners' comprehension. For example, Rosenshine and Meister (1993) studied explicit teaching before reciprocal teaching. This was the first research that tried to modify reciprocal teaching and switch it into a flexible approach that would be appropriate for students of all ages. He compared modified reciprocal teaching with traditional reciprocal teaching. Reciprocal teaching instruction occurred as a post-reading activity instead of before-reading activities. The study suggested that a modified reciprocal teaching (i.e., explicit instruction before the reciprocal teaching) could yield more significant results than reciprocal teaching only.

The goal of reciprocal teaching is to construct the meaning of the text and to check comprehension. In reciprocal teaching, the acquisition of the strategies is not the ultimate purpose of instruction. The strategies are not purposes, but a means to an end. They offer the vehicle for teaching students to read for meaning and to monitor their reading to ensure that they understand (Palincsar & Brown, 1986). Reading strategies should be taught and manipulated by instructors to enhance the efficiency of learning. However, a search of the literature revealed that many elementary school teachers have not emphasized comprehension strategy instruction in their curriculum (Pilonieta & Medina, 2009). Therefore, the present study is an attempt to bridge this gap.

## Method

**Research Design**

The overall design of this study combined both quantitative and qualitative research methods. A quasi-experimental design was employed in the study. The participants divided into two groups, the experimental group and the control group, were not chosen by random assignment. Additionally, the pilot study and pretest were conducted before the formal study began in order to control the similarities between the two groups and to examine the validity and reliability of the instruments.

The intervention lasted twelve weeks, and there were two English classes a week. Six students were interviewed in both English and Chinese after the intervention to understand their feelings about reciprocal reading and to further find out possible changes between reading attitudes and English proficiency levels after the instruction of reciprocal reading. Since the teaching time was 40 minutes per period, the experimental group was arranged to receive 20 minutes of instruction of reciprocal reading and 20 minutes of regular instruction from an English textbook in each English class session, while the control group only received a 40-minute regular instruction lesson using the English textbook and supplementary reading materials without explicit instruction.

After each treatment phase, the researchers selected and interviewed 6 students from the experimental group, three of high-score team, three of low-score team. The supplementary reading materials in the control group were those adopted from the British Council's reading resources (http://learnenglishkids.britishcouncil.org), which were the same as those in the experimental group. In other words, all the participants read the same materials. However, the control group only read the materials, but the experimental group had to predict, to answer the question, to clarify the thoughts, and to summarize those materials.

**Participants**

The study was carried out in an elementary school in Taiwan. This school is located in one of the suburban districts of Dou-nan Township. The school consisted of 500 students and the majority of them were from farming families. Two classes of fifth graders participated in this study. All the participants had learned English at school for one and half years since they were third graders.

In the experimental group, one class of 26 students, 12 males and 14 females, was selected to receive reciprocal reading intervention. Students received twenty minutes of instructions on reciprocal teaching in each English class session and another twenty minutes for regular English textbook instruction.

In the control group, 27 students, 13 males and 14 females, received traditional teacher-directed instruction. There was no reciprocal teaching instruction in the English lessons. The main material for the English class was the English textbook. Although students in the control group did not receive reciprocal teaching instruction, the researchers still provided on-line stories in class as their supplementary reading materials to each of the students for them to read.

Before implementing the study, the researchers thought it was necessary to compare the English word recognition and reading comprehension proficiency of the students in the experimental and the control groups. The researchers divided the experimental group into two teams: the high-score team and low-score team, and the same division of team were done for the control group. The number of the high-score team and low-score team was at least 10 to 15 people (Gay, Mills, & Airasian, 2009). Therefore, the researchers divided the experimental and the control groups into two teams to examine the effect of reciprocal teaching. Each high-score team and low-score team consisted of 13 or 14 students. Independent t-test (Table 1) was conducted to evaluate the hypothesis that students of high and low-score team in the experimental group ($t = -6.496$, $p < .001$) and students of high and low-score team in the control group ($t = -8.311$, $p < .001$) was found to be significant in pretest of word recognition and reading comprehension.

Table 1.
Independent sample t-test for word recognition and reading comprehension scores between high-score team and low-score team in the experimental group and the control group

| | | | | | | | 95% CI | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *n* | *M* | *SD* | *t* | *df* | *p* | LL | UL | $\eta^2$ | 1-β |
| Experimental group | | | | | | | | | | |
| High-score team | 13 | 170.54 | 21.80 | -6.496 | 24 | <.001 | -88.16 | -44.61 | .637 | 1.000 |
| Low-score team | 13 | 105.15 | 29.01 | | | | | | | |
| Control group | | | | | | | | | | |
| High-score team | 14 | 181.79 | 17.02 | -8.311 | 25 | <.001 | -87.08 | -25.49 | .734 | 1.000 |
| Low-score team | 13 | 112.00 | 26.01 | | | | | | | |

Table 2.
Independent-sample t-test results on word recognition and reading comprehension tests between-group performance in pretest

| | *n* | *M* | *SD* | *t* | *df* | *p* | 95% CI LL | 95% CI UL | $\eta^2$ | 1-β |
|---|---|---|---|---|---|---|---|---|---|---|
| All participants | | | | | | | | | | |
| Experimental group | 26 | 137.9 | 41.8 | -.904 | 51 | .370 | -33.3 | 12.2 | .016 | .144 |
| Control group | 27 | 148.2 | 41.7 | | | | | | | |
| High-score team | | | | | | | | | | |
| Experimental group | 13 | 170.5 | 21.8 | -1.501 | 25 | .145 | -26.7 | 4.2 | .083 | .303 |
| Control group | 13 | 181.8 | 17.0 | | | | | | | |
| Low-score team | | | | | | | | | | |
| Experimental group | 13 | 105.2 | 29.0 | -.634 | 24 | .532 | -29.2 | 15.5 | .016 | .093 |
| Control group | 13 | 112.0 | 26.0 | | | | | | | |

Table 2 indicated that results of all participants in experimental group and control group did not have significant difference in the pretest of word recognition and reading comprehension tests (t=-.904, p=.370). Second part of Table 2 demonstrated that results of high-score team in experimental group and control group did not have significant difference in the pretest of word recognition and reading comprehension tests (t=-1.501,p=.145). Third part of Table 2 showed that results of low-score team in experimental group and control group did not have significant difference in the pretest of word recognition and reading comprehension tests (t=-.634,p=.532).

The results illustrated that high-score and low-score teams in experimental and control group have same initial point on the basis of their English proficiency level.

**Instruments**

First, pretest and posttest were mainly adopted from Chen (2002) and partly from Hong et al. (2006). The participants were required to take the pretest as a placement-test and the posttest as an achievement assessment to evaluate their progress after 12 weeks of instruction. Second, interviews were conducted to obtain more information about their opinions and feelings about and attitudes toward reciprocal teaching. Reciprocal teaching was traditionally used with small groups independently. Due to classroom management concerns, and because the strategies would be modeled and taught during read-aloud sessions, the reciprocal teaching lessons the researchers presented were done with the whole class.

*Word Recognition Test*

The test was intended to examine students' ability of decoding words and constructing the meanings of words. The formats of the test items were partly adapted from Hong et al. (2006). There are 100 items in this test and the researchers chose all of them from students' vocabulary lists. Students had to look at the Chinese words and choose the correct English word which fit in with the meaning and write down the Chinese translation into the blank. The teacher would read aloud the sound of each item. To recognize a word, students may take advantage of several decoding skills such as morphemic clues, phonemic clues and phonic clues. After having the letter-sound correspondence ability to name a word, students could utilize this background knowledge and memory of their verbal vocabulary to identify the meaning of a word. The formats of the test items were partly adapted from Hong et al. (2006). The researchers exchanged 29 words based on English curriculum of primary level basic words list (MOE, 2008).

*Reading Comprehension Test*

The purpose of the test was aimed at examining students' understanding about simple words, sounds, word classification, and basic sentences. The formats of the test items were adopted from Chen (2002). The test was divided into four parts including word choice, sound choice, word classification, and sentence choice. There were totally 120 test items in this reading comprehension test. Students had to read each of the words and choose its corresponding pictures and English translation in 25 test items in the first part, and listen to the sound of words to choose two items to select the one correct answer in 38 test items in the second part. In the third part, there are 18 items including synonyms, antonyms, and words with the same classification. Students had to read each of the sentences and choose its corresponding pictures and English questions in 39 test items in the fourth part. The data would be analyzed according to the four parts of the reading comprehension test.

*Interviews*

The researchers conducted semi-structured interviews (Berg, 2007; Tutty, Rothery, & Grinnell, 1996). The outline of the interview questions was previously arranged. The interview questions consisted of two parts. The first part is guided questions, and the other part is for explaining for details. The researchers selected six students as the interviewees in accordance with purposeful sampling strategy (Patton, 1990).The aim of the interview is to know the experiment group students' perceptions about the curriculum and assess improvement of reading comprehension.

Table 3.
Reliability and validity of Instruments

| | Reliability | Validity |
|---|---|---|
| Word recognition test | α=.979 | *r* =.808, *p* < .001 |
| Reading comprehension test | | |
| Word choice | α=.834 | *r* =.796, *p* < .001 |
| Sound choice | α=.900 | *r* =.848, *p* < .001 |
| Word classification | α=.895 | *r* =.846, *p* < .001 |
| Sentence choice | α=.944 | *r* =.726, *p* < .001 |

**Data Collection Procedures**

Measures of pretest were administered to all of the participants to obtain their beginning level on reading comprehension and word recognition abilities and to understand their attitudes toward English reading. Both of the two groups were instructed by the researchers for 12 weeks. After the instruction for 12 weeks, both groups took a posttest of reading comprehension and word recognition to assess for possible differences of students' reading abilities before and after the reciprocal teaching program. Following the posttest, semi-structure interviews were conducted with 6 students to explore in depth about their attitude toward English reading and reciprocal teaching. Each of the interviewees' responses was transcribed.

**Data Analysis**

The software Statistical Package for Social Science (SPSS 18.0) was utilized to analyze the collected data in this study. Between-group performances in the pretest of word recognition and reading comprehension were analyzed by independent sample t-test, and paired-sample t-test was applied to compare students' performance within the experimental group and the control group. Also, the interviewees' responses were transcribed to understand students' feelings about and attitudes toward reading and reciprocal teaching.

## Results and Discussion

**Results for Research Question 1: To what extent can students perform differently in word recognition before and after the implementation of reciprocal teaching instruction?**

Table 4 demonstrated the effects of reciprocal teaching on students' word recognition ability via evaluating the improvement between the pretest-posttest within experimental and control group, high-score team between group, and low-score team between group.

Table 4.
Paired-sample t-test results of the experimental and the control groups on English word recognition test

| | | pretest | | posttest | | | | | 95% CI | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *n* | *M* | *SD* | *M* | *SD* | *t* | *df* | *p* | LL | UL |
| All participants | | | | | | | | | | |
| Experimental group | 26 | 48.42 | 23.69 | 56.12 | 19.81 | -5.691 | 25 | <.001 | -10.48 | -4.91 |
| Control group | 27 | 52.19 | 24.14 | 53.63 | 22.58 | -.741 | 26 | .456 | -5.45 | 2.56 |
| High-score team | | | | | | | | | | |

| | *n* | *M* | *SD* | *M* | *SD* | *t* | *df* | *p* | LL | UL |
|---|---|---|---|---|---|---|---|---|---|---|
| Experimental group | 13 | 66.62 | 15.43 | 71.08 | 14.89 | -3.852 | 12 | .002 | -6.99 | -1.94 |
| Control group | 14 | 71.64 | 13.84 | 72.14 | 12.18 | -.176 | 13 | .863 | -6.64 | 5.64 |
| Low-score team | | | | | | | | | | |
| Experimental group | 13 | 30.23 | 14.65 | 41.15 | 10.53 | -5.131 | 12 | <.001 | -15.56 | -6.29 |
| Control group | 13 | 31.23 | 11.74 | 33.69 | 10.63 | -.900 | 12 | .386 | -8.42 | 3.50 |

The first part of Table 4 demonstrated the results of a paired-sample t-test of pretest and posttest for the experimental group on word recognition test. It showed that participants in the experimental group made prominent progress. The development of word recognition within the experimental group reached the significant difference (t=-5.691, p<.001). However, the development of word recognition within the control group did not reach the significant difference (t=-.741, p=.465). Therefore, the statistical results implied that reciprocal teaching was facilitative to improve students' English word recognition skills.

The second part of Table 4 showed that the results of a paired-sample t-test of pretest and posttest for high score team in experimental and control group on Word Recognition Tests. It showed that participants in the experimental group made prominent progress. The development of word recognition within the experimental group reached the significant difference (t=-3.852, p=.002). However, the development of word recognition within the high score team in control group did not reach the significant difference (t=-.176, p=.863). It revealed that the students of high score team performed better on word recognition skill after receiving the instruction of reciprocal teaching.

The third part of Table 4 illustrated that the results of a paired-sample t-test of pretest and posttest for low-score team in experimental and control group on word recognition tests. It showed that participants in the experimental group made prominent progress. The development of word recognition within low-score team in the experimental group reached the significant difference (t= -5.131, p<.001). However, the development of word recognition within the low-score team in control group did not reach the significant difference (t= -.900, p=.386). It indicated that the students of low-score team also performed better on word recognition skill after receiving the instruction of reciprocal teaching.

The data showed that, after reciprocal teaching instruction, participants in experimental group made progress in word recognition skill. Also, participants in high and low score team of experimental group both increased their marks in the word recognition posttest. In control group, neither high nor low team made progress in word recognition skill without reciprocal teaching.

**Results for Research Question 2: To what extent can students perform differently in reading comprehension before and after the implementation of reciprocal teaching instruction?**

There are four sections in the reading comprehension test including word choice, sound choice, word classification, and sentence choice. Each item was analyzed by a two-tailed paired-sample t-test.

*Reading comprehension in word choice*

The first part of Table 5 demonstrated the pretest and posttest results of a paired-sample t-test for the experimental and the control group on the reading comprehension tests in word choice. The result of word choice in the experimental group was found to be significant (t =-6.692, p < .001). However, the result of word choice in the control group was not found to be significant (t = .209, p = .836). The results showed that the participants in the experimental group made significant progress in word choice after reciprocal teaching.

Table 5.
Paired-sample t-test results on the reading comprehension in word choice between- group performance

| | | pretest | | posttest | | | | | 95% CI | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *n* | *M* | *SD* | *M* | *SD* | *t* | *df* | *p* | LL | UL |
| All participants | | | | | | | | | | |
| Experimental group | 26 | 18.31 | 4.67 | 20.65 | 4.57 | -6.692 | 25 | <.001 | -3.07 | -1.62 |
| Control group | 27 | 21.26 | 4.46 | 21.19 | 4.11 | .209 | 26 | .836 | -0.65 | 0.80 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| High-score team | | | | | | | | | | |
| Experimental group | 13 | 21.23 | 1.96 | 23.38 | 1.12 | -4.779 | 12 | <.001 | -3.14 | -1.17 |
| Control group | 14 | 24.00 | 0.88 | 23.86 | 1.03 | .806 | 13 | .435 | -0.24 | 0.53 |
| Low-score team | | | | | | | | | | |
| Experimental group | 13 | 15.38 | 4.81 | 17.92 | 5.11 | -4.613 | 12 | .001 | -3.74 | -1.34 |
| Control group | 13 | 18.31 | 4.91 | 18.31 | 4.25 | .000 | 12 | 1.000 | -1.58 | 1.58 |

The second part of Table 5 indicated that the results of a paired-sample t-test of pretest and posttest for high-score team in experimental and control group on word choice of reading comprehension. It demonstrated the results of a paired-sample t-test on the reading comprehension test in word choice for the high-score team of the experimental group and the control group. The result of word choice in high-score team of the experimental group was found to be significant ($t =-.779$, $p < .001$). However, the result of word choice in high-score team of the control group was not found to be significant ($t = .806$, $p = .435$). The results showed that the participants of high-score team in the experimental group made significant progress in word choice after reciprocal teaching.

The third part of Table 5 indicated that the results of a paired-sample t-test of pretest and posttest for low-score team in experimental and control group on word choice of reading comprehension. It demonstrated the results of a paired-sample t-test on the reading comprehension tests in word choice for the low-score team of the experimental group and the control group. The result of word choice in the low-score team of the experimental group was found to be significant ($t = -4.613$, $p = .001$). However, the result of word choice in the low-score team of the control group was not found to be significant ($t = .000$, $p = 1.000$). The results demonstrated that the participants of the low-score team in the experimental group had made significant progress in word choice after reciprocal teaching.

*Reading comprehension in sound choice*

The first part of Table 6 demonstrated the pretest and posttest results of a paired-sample t-test for the experimental and the control group on the reading comprehension tests in sound choice. The result of sound choice in the experimental group was found to be significant ($t = -4.421$, $p < .001$). However, the result of sound choice in the control group was not found to be significant ($t =-.145$, $p = .886$). The results showed that the participants in the experimental group made significant progress in sound choice after reciprocal teaching.

Table 6.
Paired-sample t-test results on the reading comprehension in sound choice between- group performance

| | | pretest | | posttest | | | | | 95% CI | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *n* | *M* | *SD* | *M* | *SD* | *t* | *df* | *p* | LL | UL |
| All participants | | | | | | | | | | |
| Experimental group | 26 | 30.19 | 5.09 | 33.04 | 4.19 | -4.421 | 25 | <.001 | -1.17 | -1.52 |
| Control group | 27 | 32.59 | 3.66 | 32.67 | 4.22 | -0145 | 26 | .886 | -1.23 | 0.98 |
| High-score team | | | | | | | | | | |
| Experimental group | 13 | 33.08 | 4.59 | 35.62 | 2.06 | -2.423 | 12 | .032 | -4.82 | -0.26 |
| Control group | 14 | 34.93 | 1.82 | 35.64 | 1.60 | -1.933 | 13 | .075 | -1.51 | 0.08 |
| Low-score team | | | | | | | | | | |
| Experimental group | 13 | 27.31 | 3.86 | 30.46 | 4.24 | -4.027 | 12 | .002 | -4.86 | -1.45 |
| Control group | 13 | 30.08 | 3.50 | 29.46 | 3.80 | .634 | 12 | .538 | -1.50 | 2.73 |

The second part of Table 6 indicated that the results of a paired-sample t-test of pretest and posttest for high-score team in experimental and control group on sound choice of reading comprehension. It demonstrated the results of a paired-sample t-test on the reading comprehension test in sound choice for the high-score team of the experimental group and the control group. The result of sound choice in high-score team of the experimental group was found to be significant ($t =-2.423$, $p < .032$). However, the result of sound choice in high-score team of the control group was not found to be significant ($t = -1.933$, $p = .075$). The results showed that the participants of high-score team in the experimental group made significant progress in sound choice after reciprocal teaching.

The third part of Table 6 indicated that the results of a paired-sample t-test of pretest and posttest for low-score team in experimental and control group on sound choice of reading comprehension. It demonstrated the results of a paired-sample t-test on the reading comprehension tests in sound choice for the low-score team of the experimental group and the control group. The result of sound choice in the low-score team of the experimental group was found to be significant (t = −4.027, p = .002). However, the result of sound choice in the low-score team of the control group was not found to be significant (t = .634, p = .538). The results demonstrated that the participants of the low-score team in the experimental group had made significant progress in sound choice after reciprocal teaching

*Reading comprehension in word classification*

The first part of Table 7 demonstrated the pretest and posttest results of a paired-sample t-test for the experimental and the control group on the reading comprehension tests in word classification. The result of word classification in the experimental group was found to be significant (t = −2.289, p < .031). However, the result of word classification in the control group was not found to be significant (t =-.685, p = .499). The results showed that the participants in the experimental group made significant progress in word classification after reciprocal teaching.

The second part of Table 7 indicated that the results of a paired-sample t-test of pretest and posttest for high-score team in experimental and control group on word classification of reading comprehension. It demonstrated the results of a paired-sample t-test on the reading comprehension test in word classification for the high-score team of the experimental group and the control group. The result of word classification in high-score team of the experimental group was not found to be significant (t = −1.196, p = .255). Also, the result of word classification in high-score team of the control group was not found to be significant (t = .000, p = 1.000). The results showed that the participants of high-score team in the experimental group and control group did not make significant progress in word classification after reciprocal teaching.

The third part of Table 7 indicated that the results of a paired-sample t-test of pretest and posttest for low-score team in experimental and control group on word classification of reading comprehension. It demonstrated the results of a paired-sample t-test on the reading comprehension test in word classification for the low-score team of the experimental group and the control group. The result of word classification in low-score team of the experimental group was not found to be significant (t = −2.008, p = .086). Also, the result of word classification in low-score team of the control group was not found to be significant (t = -.879, p = .397). The results showed that the participants of low-score team in the experimental group and control group did not make significant progress in word classification after reciprocal teaching.

Table 7.
Paired-sample t-test results on the reading comprehension in word classification between-group performance

| | | pretest | | posttest | | | | | 95% CI | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *n* | *M* | *SD* | *M* | *SD* | *t* | *df* | *p* | LL | UL |
| All participants | | | | | | | | | | |
| Experimental group | 26 | 11.88 | 3.20 | 12.96 | 3.28 | -2.289 | 25 | .301 | -2.05 | -0.11 |
| Control group | 27 | 11.81 | 4.46 | 12.15 | 4.03 | -.685 | 26 | .499 | -1.33 | 0.67 |
| High-score team | | | | | | | | | | |
| Experimental group | 13 | 14.15 | 1.68 | 14.62 | 1.98 | -1.196 | 12 | .255 | -1.30 | 0.38 |
| Control group | 14 | 15.29 | 0.99 | 15.29 | 1.98 | .000 | 13 | 1.000 | -1.30 | -1.30 |
| Low-score team | | | | | | | | | | |
| Experimental group | 13 | 9.62 | 2.73 | 11.31 | 3.55 | -2.008 | 12 | .068 | -3.53 | 0.14 |
| Control group | 13 | 8.08 | 3.57 | 8.77 | 2.65 | -.879 | 12 | .397 | -2.41 | 1.02 |

*Reading comprehension in sentence choice*

The first part of Table 8 demonstrated the pretest and posttest results of a paired-sample t-test for the experimental and the control group on the reading comprehension tests in sentence choice. The result of sentence choice in the experimental group was found to be significant ($t = -5.302$, $p < .001$). However, the result of sentence choice in the control group was not found to be significant ($t = -.241$, $p = .811$). The results showed that the participants in the experimental group made significant progress in sentence choice after reciprocal teaching.

The second part of Table 8 indicated that the results of a paired-sample t-test of pretest and posttest for high-score team in experimental and control group on sentence choice of reading comprehension. It demonstrated the results of a paired-sample t-test on the reading comprehension test in sentence choice for the high-score team of the experimental group and the control group. The result of sentence choice in high-score team of the experimental group was found to be significant ($t = -3.149$, $p < .008$). However, the result of sentence choice in high-score team of the control group was not found to be significant ($t = -1.531$, $p = .150$). The results showed that the participants of high-score team in the experimental group made significant progress in sentence choice after reciprocal teaching.

The third part of Table 8 indicated that the results of a paired-sample t-test of pretest and posttest for low-score team in experimental and control group on sentence choice of reading comprehension. It demonstrated the results of a paired-sample t-test on the reading comprehension tests in sentence choice for the low-score team of the experimental group and the control group. The result of sentence choice in the low-score team of the experimental group was found to be significant ($t = -4.298$, $p = .001$). However, the result of sentence choice in the low-score team of the control group was not found to be significant ($t = .296$, $p = .772$). The results demonstrated that the participants of the low-score team in the experimental group had made significant progress in sentence choice after reciprocal teaching.

Table 8.
Paired-sample t-test results on the reading comprehension in sentence choice between-group performance

| | | pretest | | posttest | | | | | 95% CI | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *n* | *M* | *SD* | *M* | *SD* | *t* | *df* | *p* | LL | UL |
| All participants | | | | | | | | | | |
| Experimental group | 26 | 28.27 | 8.00 | 30.88 | 7.61 | -5.302 | 25 | <.001 | -3.63 | -1.60 |
| Control group | 27 | 30.33 | 8.46 | 30.56 | 8.46 | -.241 | 26 | .811 | -2.12 | 1.67 |
| High-score team | | | | | | | | | | |
| Experimental group | 13 | 33.92 | 3.28 | 36.15 | 2.19 | -3.149 | 12 | .008 | -3.77 | -0.69 |
| Control group | 14 | 35.93 | 2.24 | 36.86 | 2.35 | -1.531 | 13 | .150 | -2.24 | 0.38 |
| Low-score team | | | | | | | | | | |
| Experimental group | 13 | 22.62 | 7.28 | 25.62 | 7.47 | -4.298 | 12 | .001 | -4.52 | -1.48 |
| Control group | 13 | 24.31 | 8.41 | 23.77 | 7.27 | .296 | 12 | .772 | -3.42 | 4.50 |

**Results for Research Question 3: What are the effects of reciprocal teaching on enhancing fifth-grade students' attitudes toward English reading?**

Questions 1 to 4 were opinions about reciprocal teaching and on-line stories. All interviewees showed positive attitude toward reciprocal teaching and on-line stories. Question 5 was in quest of students' awareness of their own learning after reciprocal teaching. Five sixth interviewees have improved their word recognition and reading comprehension skills. Question 6 "Do you have any other suggestions or thoughts?" was added in case other possible feedback was left out. All interviewee suggested that more stories would be introduced in class. The interview was completed by both English and Chinese.

## Discussion

***Discussion on Research Question 1: What are the effects of reciprocal teaching on enhancing fifth-grade students' word recognition ability?***

These results were the most encouraging. In association with the reciprocal teaching program, the students in the experimental group including the high-score and the low-score teams showed significant gains in word recognition. When comparing the two groups, the researchers observed that the experimental group showed significant gains in word recognition while the control group did not. Although the control group had read the same materials as the experimental group did, the control group did not receive any teacher-guided strategies instruction.

**Discussion on Research Question 2: What are the effects of reciprocal teaching on enhancing fifth-grade students' reading comprehension ability?**

There are four sections in the reading comprehension test including word choice, sound recognition, word meaning classification, and conversation matching. The results indicated that word choice, sound choice, sentence choice reached at a significant level in the experimental group after reciprocal teaching. Word classification examined students on synonyms and antonyms meanings, which were seldom taught in elementary English lessons. The concept of morphology such as synonyms and antonyms shall be added into our lesson plans, textbooks, and annual schedule. The result of classification did not reach significant level.

**Discussion on Research Question 3: How did reciprocal teaching affect fifth-grade students' attitudes toward English reading?**

1. What do you like about reciprocal teaching? Why?
All interviewees liked reciprocal teaching (6 interviewees).
2. What do you like about on-line stories? Why?
Yes, on-line stories provided English subtitles and conversation. On-line stories are interesting and students usually look forward to reading the on-line stories (6 interviewees).
3. Which parts do you like the best of the four steps of reciprocal teaching? Why?
Prediction is the favorite part for all the interviewees, especially for slow-readers (3 interviewees from low-score team). Questioning improved vocabulary gains for faster readers (3 interviewees from high-score team). Summarization helped readers organize different sentences into short articles and increased their ability to read them out with the assistance of teacher (3 interviewees from high-score team).
4. Which parts do you like the least of the four steps of reciprocal teaching? Why?
Summarization is the most difficult part of reciprocal teaching. Three students expressed that they could not read the short summarization at once (3 interviewees from low-score).
5. After reciprocal teaching, do you feel progress in your English ability?
The interviewee discovered a way of learning, predicting what would happened next, noticing new words and connecting the meanings and sounds, and summarizing words and sentences in the story (1 interviewee from high-score team).
6. Do you have any suggestions to improve class instruction?
More stories will be introduced next semester and will have opportunities to create their own stories or imagine different story endings (6 interviewees).

## Conclusion and Implications

Before reading, the researchers attempted to activate and to build background knowledge for students. It was of importance to give an introduction, to allow students to think about the topic and to review relevant vocabulary words. By classification of words, the researchers were able to preview the new words that would appear in stories, and also review the previously learned words. The methodology, thus, connected the new and old words, and renewed the participants' wording databases. Those little reminders such as pictures or words on the blackboard seemed to awaken student's curiosity and provided scaffolding for readers. The results indicated that reciprocal teaching is beneficial to word recognition skills. As a result of using on-line stories to assist reciprocal teaching, the poorer decoders connected the meanings with word and easily memorized new words.

As to raising questions in reciprocal teaching, teachers should be aware that questions which are easy to answer with short response are better prepared for slow readers. And other questions that are open-ended and require critical thinking are provided for fast readers.

Besides, by showing students that it is all right to ask questions when they do not understand, teachers could teach them how to ask questions about stories. Hence, students' comprehension and involvement would increase.

In the present study, the researchers recommended a reading model adopted from reciprocal teaching. In reciprocal teaching, all students were firstly instructed in the four steps of prediction, question, clarification, and summarization. Students' Student's English comprehension improved after reciprocal teaching was implemented. Therefore, reciprocal teaching is worthy of future research. The students without adequate decoding skills also showed improvements in reading when they were taught reading strategies using the reciprocal teaching model.

Students felt joyful when participating in activities and imagining what might happen next. Thus, the act of motivating learners' interest may effectively spur them to actively engage in learning. Furthermore, the future researchers should create more reading activities, and attempt to provide a supportive learning atmosphere in order to help students reinforce their learning interest in learning English.

In this study, reciprocal teaching was carried out to spur students' interest in learning English. As the results indicated, reciprocal teaching instruction did in fact increase students' interest in English learning. There was still an unresolved question as follows: how shy students can be encouraged to express their feelings, opinions and answer questions. The researchers suggested that teachers prepare tangible rewards such as snacks or stationary as little gifts, and intangible rewards such as marks or warm praises to encourage these students to become more active.

## Limitations and Suggestions for Future Research

### Limitations

The findings of this study might contribute to teachers who would like to enhance students' word recognition, reading comprehension, and learning interest. Nevertheless, based on the researchers' design in the study, the findings have limitations. The reasons are listed below.

First, English textbooks and supplementary reading materials were the main resources in this study. However, the researchers chose the online stories as supplementary material and sometimes would encounter technological problems when the internet was not working or was slow. In the researchers' view, the government, academic institutions or book companies could create a reading system such as E-book for all students and English teachers. The reading system would surely be an invaluable investment in the future educational environment.

Second, there were nineteen to twenty weeks in the second semester. The researchers applied one week to the pilot study, twelve weeks to the formal study, and one week for the interviews. Three examinations and school celebrations were excluded from our research time. The research had been constrained by monthly examinations and the school anniversary celebration. These factors all decreased the time available for instruction. Finally, the short duration of the study and the limited time for reciprocal teaching sessions might have limited the ability of the students to embrace the reciprocal teaching method.

### Suggestions for Future Research

Reciprocal teaching was effective in improving fifth-graders word recognition skills and reading comprehension skills. Based on the research results, five suggestions would be offered for future researchers and English teachers in elementary schools.

First, the researchers suggested that there should be reading strategy instruction such as reciprocal teaching for fifth-graders or sixth-graders in EFL courses in Taiwan. The reciprocal teaching lessons can be done before or during reading. Hence, practitioners could utilize reciprocal teaching flexibly and modify them to suit the classroom situation.

Second, owing to the lack of basic skills of reading and speaking, without technological support, students with low- or intermediate-level abilities may have difficulty understanding the stories and new words. Based on the collected interview data, most of the participants started learning more new English words from the on-line stories and reciprocal teaching. Hence, the assistance of on-line stories and teachers' guidance are highly recommended. Those materials should be added into our textbooks and thus enliven our English lessons. From the researchers' point of view, the government, academic institutions, or even book companies should establish an on-line story bank for all levels of readers for future English education.
Third, the reading teacher played a role of instructor, consultant, and cheerleader, who provided enthusiasm and a positive image toward reading. In a pleasant atmosphere, the students read as much as possible, perhaps in and definitely out of the classroom. The researchers suggested that reciprocal teaching be associated with the administration's economic and technical support. If possible, it would be even better to have computers for all the students, and then to let them select what on-line stories they would like to read after class. This would make reading have its own reward.

Fourth, from the researchers' perspective, reciprocal teaching in Taiwan would be modified as teacher-guided instruction in the beginning because students need time to be familiar with reciprocal teaching. The procedure of four steps of reciprocal teaching would be flexibly managed in lesson plans as well. In addition, teacher-guided instruction would better be gradually replaced by student-centered discussion after students get familiar with the procedure of reciprocal teaching in the future.

Finally, this study provided a keen insight into comprehension development of students in the researchers' classroom. The results of the study, however, may not yet be conclusive because the number of students (N=53) included in this study was limited. A similar study with more participants is suggested in the future.

## References


Alfassi, M. (1998). Reading for meaning: The efficacy of reciprocal teaching in fostering reading comprehension in high school students in remedial reading class. *American Educational Research Journal*, *35* (2), 309-332.

Berg, B. L. (2007). *Qualitative research methods for the social sciences* (6th ed.). Boston, MA: Allyn and Bacon.

Chen, T. C. (2002). *The development of English reading diagnostic test for junior high school students*. Unpublished master's thesis, National Changhua University of Education, Taiwan.

Duffy, G. G., & Roehler, L. R. (1987). Improving reading instruction through the use of responsive elaboration. *The Reading Teacher*, *40* (6), 514-519.

Fevre, D. M. L., Moore, D. W., & Wilkinson, I. A. G. (2003). Tape-assisted reciprocal teaching: Cognitive bootstrapping for poor decoders. *British Journal of Educational Psychology*, *73* (1), 37-58. doi:10.1348/000709903762869905

Fielding, L., & Roller, C. (1992). Making difficult books accessible and easy books acceptable. *The Reading Teacher*, *45* (9), 678-685.

Gay, L. R., Mills, G. E., & Airasian, P. (2009). *Educational research: Competencies for analysis and applications* (9th). Upper Saddle River, NJ: Prentice Hall.

Hong, Y. L., Huang, H. S., Lin, C. R., Chou, Y. L., Liou, Y.-M., & Hsie, L. H. (2006). The development of the Englsih word recognition test for junior high and elementary school students. *Psychological Testing*, *53* (2), 155-180.

Jenkins, J. R., Stein, M. L., & Wysocki, K. (1984). Learning vocabulary through reading. *American Educational Research Journal*, *21* (4), 767-787.

Ministry of Education. (2008). *Guidelines of English curriculum of grade 1st-9th*. Taiwan: MOE.

Myers, P. A. (2006). The princess storyteller, clara clarifier, quincy questioner, and the wizard: Reciprocal teaching adapted for kindergarten students. *The Reading Teacher*, *59* (4), 314-324.

Neuman, W. L. (2006). *Social research methods: Qualitative and quantitative approaches* (6th ed.). Boston: Pearson.

Oczkus, L. D. (2003). *Reciprocal teaching at work: Strategies for improving reading comprehension* (2nd ed.). Newark, DE: International Reading Association.

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension fostering and comprehension monitoring activities. *Cognition and Instruction*, *1* (2), 117-175. doi:10.1207/s1532690xci0102_1

Palincsar, A. S., & Brown, A. L. (1986). Interactive teaching to promote independent learning from text. *The Reading Teacher*, *39* (8), 771-777.

Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2nd ed.). Newbury Park, CA: Sage.

Pilonieta, P., & Medina, A. L. (2009). Reciprocal teaching for the primary grades: We can do it, too! *The Reading Teacher*, *63* (2), 120-129. doi:10.1598/RT.63.2.3

Rosenshine, B., & Meister, C. E. (1993, May). *Reciprocal teaching: A review of 19 experimental studies* (tech. rep. No. 574). Urbana, IL. Retrieved from https://www.ideals.illinois.edu/bitstream/handle/2142/17744/ctrstreadtechrepv01993i00574_opt.pdf?sequence=1

Tutty, L. M., Rothery, M., & Grinnell, R. M. (1996). *Qualitative research for social workers: Phases, steps, and tasks.* London: Allyn and Bacon.


International Journal of Contemporary Educational Research (IJCER)

**www.ijcer.net**

# National Mathematical Centre – Mathematics Improvement Project (NMC-MIP): A Project Transforming The Mathematics Performance Of Students

**Lawal. O. Adetula**[1]

[1] National Mathematical Centre, Abuja

**To cite this article:**

Adetula, O. L. (2015). Mathematics improvement project (NMC-MIP): a project transforming the mathematics performance of students. *International Journal of Contemporary Educational Research, 2*(2), 104-117.

IJCER International Journal of Contemporary Educational Research
Volume 2, Number 2, July 2015, Page 104-117 ISSN: 2148-3868

# National Mathematical Centre – Mathematics Improvement Project (NMC-MIP): A Project Transforming the Mathematics Performance of Students

**Lawal. O. Adetula**[1*]
[1] National Mathematical Centre, Abuja

## Abstract

The historical perspective of the deplorable mathematics West African Senior Secondary Certificate Examination (WASSCE) performance in the Nigerian secondary schools is explicated upon. National Mathematical Centre (NMC) created Mathematics Improvement Project (MIP), and the project produced germane instructional materials as tools and weapons to banish all the weak and faulty mathematical foundations in MIP (experimental) schools. To achieve this, the NMC resource personnel relying on these instructional materials, trained the MIP teachers and students to effectively use active or model teaching methods to teach and learn mathematics concepts in the curriculum.

The impacts of this project (MIP) on students and teachers taken part are tremendous and fulfilling. The mathematics WASSCE high success rate in project (MIP) schools exemplified this. This is even more so (1) using descriptive statistics when comparison is made between mathematics WASSCE nationwide performances before and after the inception of MIP, (2) using non-parametric statistics when comparison is made between mathematics WASSCE performances for MIP and non MIP students in Kogi State.

**Key words:** MIP sermon, instructional system, model teaching, project sustainability, electronic classroom mechanism, significance of difference between two proportions.

## Introduction

> *Mathematics is the gate and key of the sciences. Neglect of Mathematics works injury to all knowledge, since he who is ignorant of it cannot know the other sciences or the things of this world. And what is worse, men who are thus ignorant are unable to perceive their own ignorance, and so do not seek a remedy.*
>
> *Kline (1972) quoting Roger Bacon.*

Today, Mathematics is the linchpin in the task of national capacity building in science and technology and therefore, any shortcomings in this subject constitute drawbacks to the achievement of our science and technology objectives or rather, the Millennium Development Goals (MDGs) and the transformation agenda (Adetula, 2010; Ugoh, 1980). This remark was precisely synonymous to what Fafunwa (1995), former Minister of Education predicted some years ago, when he succinctly professed: "science and technology have become a dominant cultural factor, and any nation who is not alive to this fact is either dead or dying"; And as an appendage, it can be quickly added that "mathematics is the best cognitive tool that moves these science and technology activities in today's technological globalization".

This statement is true because without the knowledge of mathematics (head of all knowledge that often gives highly impressive accomplishments), there is no science, without science there is no modern technology, and without modern technology there is no modern society (Ale and Adetula, 2010). The message is clear. The prosperity of any country and the development of mathematical sciences are intimately connected.

* Corresponding Author: *Lawal. O. Adetula, lawaladetula@yahoo.com*

**Historical Background**

One undisputable or irrefutable fact is that the students' WASCE/ WASSCE poor performances throughout the Federation over the years in this subject called mathematics are reflection, at least in part, of the poor state of teaching and learning processes in the Nigerian schools and a serious indictment on secondary education in Nigeria.

The two data below speak volumes in regard of the sorry state of the performances of students nationwide over the years in mathematics in West African School Certificate Examination (WASCE, prior to 1986), and West African Senior School Certificate Examination (WASSCE, after 1986).

The first data depicted below is the sample of the yearly average mathematics performance of students nationwide in (WASCE) from the inception of secondary schools in Nigeria until 1985, (note, the results prior to 1964 are not significantly different). During this period, mathematics was compartmentalized into arithmetic, algebra processes, geometry, trigonometry and statistics/ probability. That is, each of the five branches of mathematics is considered a closed and separate field of teaching and learning. This is the classical view point of mathematics.

| **Year** | **1964** | **1967** | **1969** | **1981** | **1982** | **1983** | **1984** | **1985** |
|---|---|---|---|---|---|---|---|---|
| % pass (at Credit level | 6.9* | 7.8* | 8.59 | 6.26 | 13.30 | 9.42 | 11.26 | 11.45 |

Source: WAEC, except * that are from secondary source.

The performance of students in mathematics WASCE during classical view point of mathematics was nothing to write home about; hence a contemporary view point of mathematics came into being, following the Benin Mathematics Conference in 1976. Actually, this was the effect of *sputnik* and the ripples and waves on *modern mathematics* that generated all over the world then (Adetula, 1981). The conference resolution recognized that certain fundamental concepts (e.g. set relations and mappings) underpinned all the branches of mathematics, and that structural concepts gave possibilities for organizing all mathematics into a unified body of knowledge. It was claimed that this structuring permits greater understanding and efficiency in learning, and uncovers concepts and theories previously hidden by the traditional separation. For this reason, a unified secondary school curriculum (three years each for junior and senior secondary) was adopted, and National Mathematics Curriculum Guides for schools (based on contemporary view point) were released in March 1979 by Federal Ministry of Education. Effective usage of these curriculum guides by many Junior Secondary Schools started in 1981 (Adetula, 1989), and therefore Senior Secondary School started in 1984. Apart from putting in place a unified curriculum, the effort of Mathematics Association of Nigeria (MAN) to assist mathematics teachers (by conducting teacher training mathematics workshops) on how to teach the contents of this unified curriculum must be applauded, and the slogan *war against poor performance in mathematics* was declared. Again, notwithstanding all these efforts, the performances of students nationwide in mathematics in West African Senior Secondary Certificate Examination (WASSCE), beginning from 1986 were as poor as the result of WASCE under classical view point of mathematics. Sample of these performances are depicted in the second data below.

| Year | 1986 | 1989 | 1991 | 1997 | 1998 | 1999 |
|---|---|---|---|---|---|---|
| % pass (at Credit level | 10.83 | 8.78 | 11.13 | 7.62 | 10.91 | 18.07 |

Source: West African Examination Council (WAEC), 2000.

The National Mathematical Centre (NMC) that was established in 1988 and started functioning in 1989 was not ignorant of the depressing mathematics situation in Nigerian schools. The Centre has a mandate among others to promote the power and potentials of mathematics among the secondary school teachers and students, which will in turn enhance the social, commercial, scientific, industrial and technological development of the nation. One slogan the Centre shares very much is that '*Any nation that treats the mathematics education of her youths with levity does so at her own scientific and technological peril'*. Therefore, to achieve this particular mandate, the Centre in 2002, embarked on a project - *Mathematics Improvement Project (MIP)* as a panacea to remedy the appalling mathematics situation in the schools. The author, who played a leading role in the activities during the *war against poor performance in mathematics,* was given a job as a Resident Consultant (MIP) to develop a proposal for NMC Mathematics Improvement Project, coordinate the project implementation throughout the country and also write research reports to cover the project evaluation activities.

NMC-MIP, like any other Mathematics Improvement Project: Chicago, Michigan, North Carolina, Florida, Iowa City, Johannesburg, (Larmer, 2010; Ulm, 2011; Deubel, 2014) is a research study on the improvement of depressing mathematics teaching and learning at senior secondary schools. Therefore, the Project's main goal is to use the model of Instructional System designed by Romberg (1968), and modified for MIP by Adetula (2002), to promote (through workshops' activities within two years) the roles of mathematics teachers (class teaching, student assessment and managerial activities) as well as the roles of students to learn mathematics with ease. *This will translate to students' greater performance measurable by the number of credit passes in the public mathematics examinations (e.g. WASSCE).*

The other purpose of this paper is on contemporary experience sharing in project research management, where other countries in the world can benefit after reading this report. Finally, one best way in carrying out a good research lies in an effective combination (hybrid) of qualitative (descriptive) and quantitative techniques exhibited in this paper.

***Model: Basic Components of MIP Instructional System***

For clarity sake, the model of MIP Instructional System depicted in Fig 1 below is used as a framework for the examination of NMC- MIP.

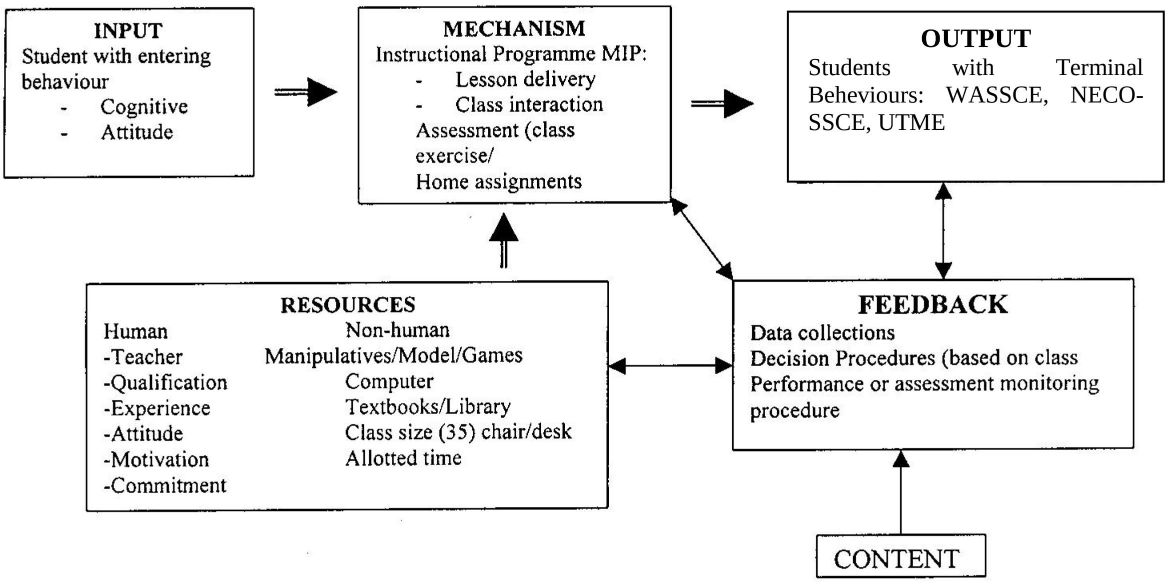



Fig 1: Model of MIP Instructional System.

The word system here refers to a 'man-made, controlled, functional structure'.

- A man-made structure is one that has interdependent components that can be changed or manipulated
- Controlled means that there is a feedback or monitoring procedure that can be used to manage the system.
- Functional means that the system is goal-oriented with a stated purpose or intent.

This system has five basic components: Input, Output, Mechanism, Feedback and Resources. Though the focus of this instructional system is not on content (since the competence of Nigerian Educational Research and Development Council -NERDC-, that specializes in producing *National Mathematics Curriculum Guides* is never in doubt), but it must be noted that the content is shown in the model leading to the feedback element; this implies that content helps define the boundary conditions within which the system operates.

The MIP instructional system has students as input; here, the current state of the art for each student as regards his/her cognitive and attitudinal behaviours can be determined using questionnaires on entering behaviours (Adetula, 2013a). These same students with changed behaviour (after going through the instructional and awareness workshops, i.e. after the intervention) became the output. The instructional programme that relied on training workshops and effective classroom lessons is explicated under the description of MIP teaching methodology exemplified in model teaching method. Actually, description of this instructional methodology encompasses the mechanism or instructional program (lesson delivery and class interaction), and feedback (assessment and monitoring procedures).

The resources which are human (e.g. mathematics teachers), and non-human (e.g. NMC instructional materials: manipulative models, games, computers, textbooks, charts, workbooks, instructional and curriculum guidelines) make instruction easy, interesting and meaningful.

## Method

### Subjects

The subjects for the study were drawn from one major source. All second year senior secondary - SS2 students and the teachers of schools selected for MIP project by the interested State Governments in Nigeria that keyed into the project (but at different period). These States and periods are Katsina (2002/2003 and 2003/2004) academic years; Zamfara (2004/2005 and 2005/2006); Ondo (batch I) (2006/2007 and 2007/2008); Ondo (batch II) (2008/2009 and 2009/2010); And Kogi (2009/2010 and 2010/2011) academic years

For the project, each State Government selected six schools (made up of two schools in each of the three senatorial districts in the State for wider coverage), targeting in each school, SS2 students at the academic year inception of the project. The study followed the target students to SS3 in the following academic year until they sat for their WASSCE. Therefore, the project is a two academic year longitudinal intervention.

The data of MIP students and mathematics teachers in the four States is depicted on Table 1.

***Table 1**: Data of MIP Students and Mathematics Teachers in Four States.*

| | **Requirements** | **Katsina** | **Zamfara** | **Ondo** | | **Kogi** |
|---|---|---|---|---|---|---|
| | | | | Batch 1 | Batch 2 | |
| **(i)** | No of students in MIP schools | 5380 (1403) | 5010 (1298) | 2481 (716) | 2762 (792) | 3128 (1081) |
| **(ii)** | Total No of teachers in MIP schools | 34 | 33 | 17 | 19 | 12 |
| **(iii)** | Total No of qualified MIP teachers | 9 | 4 | 10 | 11 | 9 |
| **(iv)** | Students/Qualified teacher ratio | 598:1 | 1253:1 | 248:1 | 251:1 | 348:1 |
| **(v)** | Average No of students/ class | 86 | 104 | 50 | 54 | 53 |

- The data here is for the ‘pulled’ six selected senior secondary schools per State
- Number in parenthesis is the project students

### Procedure

Two major procedural activities in MIP (experimental) schools can be identified thus:

First, the SS2 students’ strengths and weaknesses in mathematics (skill and content domains) are determined together with the school mathematics staff strength and their difficult areas to teach. In addition WASSCE previous years’ results are all collated. This constitutes the base line and the needs assessment.

Secondly, These difficult topics in mathematics are addressed using student/ teacher training mathematics workshops and effective classroom lessons that often rely on NMC expertise and germane instructional materials.
Clearly, each workshop is believed to re-engineer and to reformate the MIP teacher’s roles to demonstrate such characteristics as:

- A deeper and more flexible content knowledge in the curriculum.
- Commitment to students and their learning to transform those students into mathematical dynamo.

- A wide repertoire of mathematics teaching strategies e.g. questioning techniques leading to active/ model teaching with a focus on meaning and class interaction; And the use of relevant instructional materials like NMC mathematics kits which are sources of valuable insight and intuition for the students.
- Class interaction that stresses teacher-student communication that promotes or builds student confidence, interest, self concept, and persistence.
- An effective performance assessment and portfolio evaluation technique, e.g. evaluation technique that emphasized logical process instead of product.

It is generally believed that the teaching and learning of mathematics are complex enterprises and that one of the best ways out is exemplified in active or model method of teaching/ learning mathematical concept with the view of eliciting meaning and gaining more understanding in the concept. Now let us consider one each of such active/ model teaching for MIP students in trigonometry, algebra, and geometry. These examples as highlighted below for MIP teaching episodes are merely representative and not exhaustive.

Two major procedural activities in MIP (experimental) schools can be identified thus:

First, the SS2 students' strengths and weaknesses in mathematics (skill and content domains) are determined together with the school mathematics staff strength and their difficult areas to teach. In addition WASSCE previous years' results are all collated. This constitutes the base line and the needs assessment.

Secondly, these difficult topics in mathematics are addressed using student/ teacher training mathematics workshops and effective classroom lessons that often rely on NMC expertise and germane instructional materials.
Clearly, each workshop is believed to re-engineer and to reformate the MIP teacher's roles to demonstrate such characteristics as:

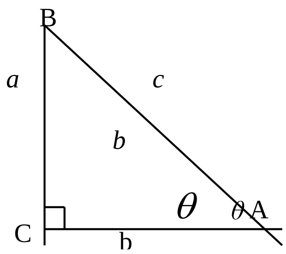
- A deeper and more flexible content knowledge in the curriculum.
- Commitment to students and their learning to transform those students into mathematical dynamo.
- A wide repertoire of mathematics teaching strategies e.g. questioning techniques leading to active/ model teaching with a focus on meaning and class interaction; And the use of relevant instructional materials like NMC mathematics kits which are sources of valuable insight and intuition for the students.
- Class interaction that stresses teacher-student communication that promotes or builds student confidence, interest, self-concept, and persistence.
- An effective performance assessment and portfolio evaluation technique, e.g. evaluation technique that emphasized logical process instead of product.

It is generally believed that the teaching and learning of mathematics are complex enterprises and that one of the best ways out is exemplified in active or model method of teaching/ learning mathematical concept with the view of eliciting meaning and gaining more understanding in the concept. Now let us consider one each of such active/ model teaching for MIP students in trigonometry, algebra, and geometry. These examples as highlighted below for MIP teaching episodes are merely representative and not exhaustive.

*Model Method of Teaching Special or Quadrantal Angles ($0^o$, $90^o$, $180^o$, $270^o$) of Trigonometrical Functions*

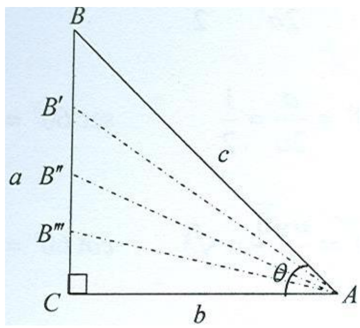
In a right angle $\Delta ABC$

$$\cos\theta = \tfrac{b}{c} \quad \sec\theta = \tfrac{c}{b}$$
$$\sin\theta = \tfrac{a}{c} \quad \operatorname{cosec}\theta = \tfrac{c}{a}$$
$$\tan\theta = \tfrac{a}{b} \quad \cot\theta = \tfrac{b}{a}$$




Will we still have a right angle $\Delta ABC$ if $\theta$ = $0^o$? or $\theta$ = $90^o$?

In the right angle triangle ABC, consider what happens when $\theta$ started decreasing. First point *B* moves to point $B'$.



Then to point $B''$ , … and finally, as point $B$

tends to point $C$, i.e., as $\theta \to 0^o$.
$\Rightarrow a \to 0$ and $c \to b$.

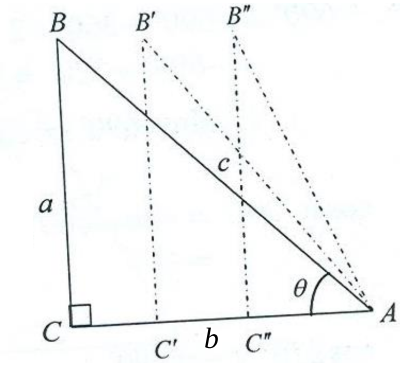
Hence,

$\sin 0^o = \frac{0}{b} = 0$ $\quad$ $cosec\, 0^o = \frac{b}{0} = \infty$

$\cos 0^o = \frac{b}{b} = 1$ $\quad$ $\sec 0^o = \frac{b}{b} = 1$

$\tan 0^o = \frac{o}{b} = 0$ $\quad$ $\cot 0^o = \frac{b}{o} = \infty$

In the right-angled $\Delta ABC$, consider what happens when $\theta$ started increasing. First point $C$ moves to $C'$ as $B$ moves to $B'$, then to $C''$ as $B'$ moves to $B''$. Finally, as $C$ tends to point $A$, $b \to 0$, $\theta \to 90^o$ and $a \to c$.




Hence

$\sin 90^0 = \frac{c}{c} = 1$ $\quad$ $\cos ec 90^0 = \frac{c}{c} = 1$

$\cos 90^0 = \frac{o}{c} = 0$ $\quad$ $\sec 90^0 = \frac{c}{o} = \infty$

$\tan 90^o = \frac{c}{o} = \infty$ $\quad$ $\cot 90^0 = \frac{0}{c} = 0$

Strictly speaking, angles 0º and 90º cannot appear in a right angled triangle as acute angles, but in an extended view of trigonometric functions, we just have to consider their values. (Note that the foundation of the concept of limit is been cultivated here). The symbol $\infty$ (infinity, which is not a number) indicates that the absolute value of the quantity increases without bound.

*Model Method of teaching the concept of the Difference of two Squares (An algebraic identity).*

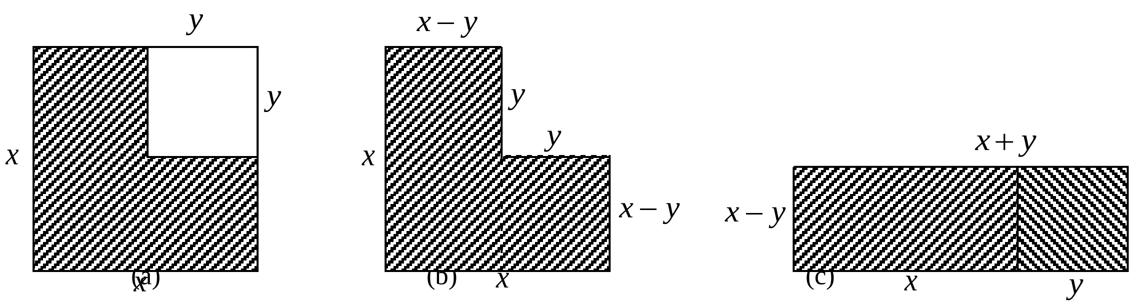
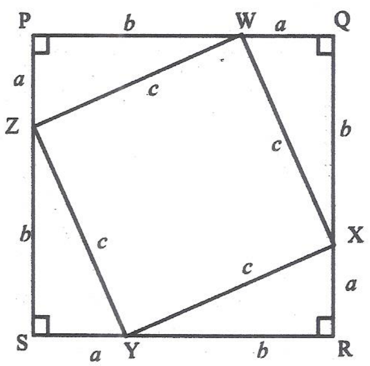



Fig. 2

In the above model (fig. 2), a square whose area is designated by $y^2$ is placed on a square whose area is $x^2$, so that the shaded portion shown in (a) is a representation of $x^2 - y^2$. The shaded region is subdivided as shown in (b) and re-arranged as in (c) to illustrate that $x^2 - y^2 = (x - y)(x + y)$. This shows clearly that the model not only promote but can be used to enrich learning.

*Model Method of Teaching Pythagoras Theorem. (A geometric identity).*

Proving the Pythagoras Theorem invoking the Chinese proof is very ingenious.




Area *WXYZ* $= c^2$
Area *PQRS* $= (a + b)(a + b) = a^2 + b^2 + 2ab$. …… (i)

But Area *PQRS* = Area of *WXYZ* + Area of 4 triangles
$= c^2 + 4 \times \frac{1}{2} ab$
$= c^2 + 2ab$ …… (ii)

Equating (i) and (ii) to get

$$c^2 + 2ab = a^2 + b^2 + 2ab$$

i.e. $c^2 = a^2 + b^2$

The suggested active method of teaching enable students to discover the truths for themselves before formal proof of propositions is introduced. Quite often these formal proof of proposition normally fogs students' mind without active method of teaching. It must be clearly stated here that students (especially non-MIP) are given these trigonometric functions and fundamental identities (algebraic and geometric) as definition to memorize without any meaning. In fact their memorization of these functions, which is synonymous to rote-drill and practice (parroted and catechetical procedures), did little to promote conceptual efficiency; instead dampen and deaden the students' enthusiasm, and also retention of their mathematical learning. The story of Bertrand Russell on this issue is fascinating. Bertrand Russell said 'that the beginning of algebra, he found far more difficult than geometry, perhaps as a result of bad teaching. He was made to learn by heart *the square of the sum of two numbers is equal to the sum of their squares increased by twice their product.* He said, he had not the vaguest idea of what this meant, and when he could not remember the words, his teacher threw the book at his head, but did not stimulate his intellect in any way.' (Russell, 1968, p. 31)

From the foregoing, it is discovered that MIP *lesson delivery* aims at presenting a clear and structured unique active/ model teaching methodology, which promotes and supports the classroom spirit of inquiry and problem solving (Adetula, 1988; 2010; & 2012). It is this inquiry that truly drives the students' learning, not a teacher telling them (the students). In addition, effective monitoring and various internal evaluation activities carried out are to assure that students are not classified to be mathematically empty (Adetula, 2013a & 2013b).

At this juncture, it must not be misconstrued that using a teaching model is the only magic bullet for mathematical success. According to MIP sermon (Adetula, 2005), it is clearly pointed out that the other dimension apart from active teaching is to build into the students the awareness that they can succeed in mathematics once they believe in themselves to succeed. MIP sermon also dwelt on the power, value, and utility of mathematics during a session in the MIP workshop. This stress the smooth applicability of mathematics to the real world, so as to demystify a number of stereotypical attitudes students have towards mathematics. The summary of MIP sermon is given below.

*The mathematical sermon for MIP students.*

This is the spiritual or awareness dimension of the project working on the students' mindset. That is, the sermon works on student's thought processes to believe in self as an achiever that can succeed in mathematics through the power of almighty God and hard work. The slogan is: '*You may succeed if nobody else believes in you, but you will never succeed if you don't believe in yourself*'. Simply put, this is a sermon that:

inspires student to mathematics success,
energizes him (or her) for achievement,
liberates him (or her) from career limitations,
frees him (or her) from mathematical fear,
brightens up his (or her) pathway, and
makes his (or her) dreams/ aspirations come true.

The prayer that follows by each student is simple, and thus:

O Lord, let me be able to fulfill my mathematical learning roles, so that the mathematical light as detailed in this sermon will shine upon alter of my life.

**Research Instruments**

The research instruments used national examinations. That is, what are considered in this project are *external evaluation activities,* like national examinations. Unlike test instruments constructed by a researcher, the validity and reliability of these national examinations produce indubitable and irrefutable conclusions and cannot be concocted.

Actually, performance of SS3 target students on Senior School Certificate Examination (SSCE) conducted by Examination Councils like West African Examination Council (WAEC) and National Examination Council (NECO), and also performance of these students on other public placement examination like Unified Tertiary Matriculation Examination (UTME) to higher institutions conducted by Examination Board like Joint Admission and Matriculation Board (JAMB) always determine in no small measure the positive effect of MIP. However, the analysis of MIP results in this paper focused on mathematics WASSCE.

## Results

*Record of WASSCE Achievement in MIP States: By Descriptive Statistics*

The impacts of this project (MIP) on students and teachers taking part are tremendous and fulfilling. After the two year intervention programme, sufficient number (at least, 90%) of MIP school teachers, inclusive of all the teachers teaching project students, have been trained during MIP workshops on how to use the NMC instructional materials as tools, weapons and arsenals to banish the weak and faulty mathematics foundation in project schools, and thereby transforming the students to become mathematical dynamos.

Actually, it is considered at this point that if the students are to be brought to the frontiers of mathematical knowledge, then the prime role of the teacher is to determine the whereabouts of these frontiers which often reside in the NMC germane instructional materials. Clearly, this role, the teachers have carried out efficiently as indicated below in the mathematics results of MIP students. These results are exemplified by the percentage of at least credit passes in the public examinations (WASSCE), and the following data (Adetula, 2012), solidify and concretize this point.

In Katsina State the average performance of six MIP pilot schools in Mathematics WASSCE in 2004 was 47%. Before MIP intervention, it was 3.8%.
In Zamfara State in 2006, it was 76%, and 14% before the intervention.
In Ondo State (Batch 1) in 2008, it was 77%, and 41% before the intervention.
In Ondo State (Batch 2) in 2009, it was 81%, and 48%, before the intervention.
And in Kogi State in 2011, it was 72%, and 19% before the intervention.

For clarity sake, MIP Kogi State detailed results will be used for illumination. These results are indicated below in Tables 2.
The data displayed on Table 2 indicated the success rate in mathematics WASSCE in the MIP schools in Kogi State. The analysis of the results revealed that the average performance (in at least credit pass) of six MIP schools in mathematics WASSCE is 72%. However, when pass level (D7 & E8) is also considered, the average total pass is 90%.

Table 2. *Comparative Analysis in Mathematics SSCE Results (at least Credit pass) in the six MIP-Pilot Schools in Kogi State*

| | | Year 2009 Mathematics Results before the Commencement of MIP Project | | Year 2010 Mathematics Results (MIP Project did not involve these students) | | Year 2011 Mathematics Result after Two Years of MIP Project | |
|---|---|---|---|---|---|---|---|
| S/N | Name of School | No. of Candidate | $A_1$ - $C_6$ % Pass | No. of Candidate | $A_1$ - $C_6$ % Pass | No of Candidate | $A_1$ - $C_6$ % Pass |
| 1 | Govt Sc Sec Sch, Icheke-Ogane (GSSS, Icheke-Ogane) | 36 | 0.00 | 26 | 7.69 | 56 | 55.56 |
| 2 | St. Peter's Coll Idah | 244 | 54.10 | 212 | 33.02 | 175 | 70.86 |
| 3 | Ochaja Boys Sc Sec Sch. Ochaja (OBSSS Ochaja) | 76 | 0.00 | 86 | No Results | 88 | 97.75 |
| 4. | Abdul-Aziz Atta Memorial Coll, Okene (AAMC, Okene) | 183 | 24.04 | 179 | 94.41 | 299 | 92.14 |
| 5. | Govt Sc Sec Sch. Ogaminana (GSSS Ogaminana) | 239 | 12.55 | 273 | 1.83 | 181 | 53.04 |
| 6. | Govt. Sc. Sec Sch. Lokoja | 218 | 9.63 | 215 | 11.16 | 282 | 57.09 |

The MIP project for the State started in year 2010 with SS2, the project was continued with these students in SS3 in year 2011. The average mathematics WASSCE result (at least credit pass) in the six MIP schools in 2011 (with total population 1081) is 72%.

To sustain the project in these schools, the project materials were retrieved from the MIP graduating set of 2011, and given to the immediate incoming set into SS3. Since the school mathematics teachers have been trained effectively to use these project materials and the State MIP Desk Officer (NMC officer who is a mathematics educator residing in the State to promote MIP activities) is still qualify to continue in organizing MIP workshops for these incoming students into SS3, then it is reasonable to conclude that the graduating students of the set 2012 will also perform well in mathematics WASSCE. Table 3 buttressed these facts. Clearly, the sustainability of the project is guaranteed through this process.

Table 3. *2012 Mathematics WASSCE Result: A Follow – up Effort for MIP Schools in Kogi State*

Year 2012 Mathematics WASSCE Results.

| **S/N** | Name of School | No of Candidate | $A_1$ - $C_6$ % Pass |
|---|---|---|---|
| 1. | GSSS, Icheke-Ogane | 96 | 70.83 |
| 2. | St. Peter's College Idah | 191 | 84.29 |
| 3. | OBSSS, Ochaja | 48 | 22.92 |
| 4. | AAMC, Okene | 121 | 80.17 |
| 5. | GSSS, Ogaminana | 116 | 96.55 |
| 6. | GSSS, Lokoja | 282 | 3.55 |

The poor performances of OBSSS, Ochaja and GSSS, Lokoja were as a result of sudden transfer of qualified mathematics teachers from these schools without immediate replacement. The message to be drawn from this is that without qualified teachers, the result of any school project will always be spurious and disastrous. Without GSSS Lokoja in the analysis of the result, the average mathematics WASSCE performance at credit level of the remaining five schools is 78.5%. This is really a good performance. With all these good performances, Is it save to claim that MIP is really preparing students for a smooth pilgrimage to advanced mathematics studies?

*Effects of MIP on mathematics WASSCE performance nationwide*.

The data below indicated the average national mathematics WASSCE results as from the year 2002 when MIP was established and various NMC sponsored national mathematics activities at senior secondary education level were vigorously pursued.

| **Year** | **2002** | **2004** | **2005** | **2006** | **2007** | **2008** | **2009** | **2012** |
|---|---|---|---|---|---|---|---|---|
| % pass at Credit level | 34.41 | 34.51 | 38.23 | 41.54 | 46.90 | 56.96 | 45.33. | 48.88 |

Source: WAEC Correspondence with National Mathematical Centre (2012),

In actual fact, there is a clear difference in these national mathematics WASSCE results as depicted above and the previous nationwide results earlier given in the background section of this paper.

The difference in performance to the previous performances before year 2002 is crystal clear. Actually, this significant improvement in performance as from the year 2002 is based on the commitment of the staff of the project using basic components of MIP instructional system and the effective MIP instructional materials that simplify, elementarise and even trivialize the mathematics concepts that students and teachers perceived to be difficult. In addition, it is also true that MIP staff vigorously pursued mathematics teacher training workshop using MIP instructional materials across many States in addition to MIP States. Therefore it is reasonable to conclude that MIP promotes national improvement in mathematics performance by using the active teaching style to attract and retain the attention of the students being taught. The MIP key words here are: attention, attraction, and retention.

**By Inferential Statistics**

It is necessary to determine if MIP students in Kogi State significantly performed better than the non-MIP students in mathematics WASSCE.
To carry out this, a purposive sample of six non-MIP schools were identified that best match the six MIP schools: in qualified teacher/ student ratio, and percentage credit performance in mathematics WASSCE before the intervention. This is to meet the assumption that the students in these two groups (MIP and non-MIP) are equivalent before the intervention as indicated on Table 4. Actually, it is reasonable to assume that purposive sampling is the best option, since there is no random sampling for the selection of MIP schools. The State Government just gave six schools by fiat (as MIP schools) for the project.

Table 4. *Comparison of Mathematics WASSCE results of MIP and non-MIP students in Kogi State using test of significance of a difference between two proportions*

| S/N | Matching a) MIP and b) Non MIP Schools | Matching results in both MIP and non -MIP Schools prior to the intervention | | | | Results: Comparing MIP schools with intervention and non MIP schools without intervention (2011) | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | 2009 | | 2010 | | | | | |
| | | N | n | N | n | N | n | Z | Decision $\alpha$ =0.01 |
| 1 | a) GSS Icheke-Ogane | 36 | 0 (0%) | 26 | 2 (7.6%) | 56 | 31 | 2.64 | Sig. |
| | b) GSS Egume | 222 | 2 (0.9%) | 247 | 14 (5.7%) | 256 | 97 | | |
| 2 | a) St. Peter's Coll. Idah | 244 | 132 (54%) | 212 | 70 (33%) | 175 | 124 | 9.76 | Sig |
| | b) CSS Adavi_eba | 314 | 178 (56%) | 332 | 113 (34%) | 307 | 14 | | |
| 3 | a) OBSSS Ochaja | 76 | 0 (0%) | 86 | 0(0%) | 88 | 86 | 6.86 | Sig |
| | b) GSS Sarki-Noma | 57 | 1(1.7%) | 58 | 0(0%) | 40 | 20 | | |
| 4 | a) AAMC Okene | 183 | 44 (24%) | 179 | 169 (94%) | 299 | 25 | 23.0 | Sig |
| | b) ECSS Ogaminana | 235 | 88 (37%) | 223 | 152 (68%) | 362 | 12 | | |
| 5 | a) GSSS Ogaminana | 239 | 30 (12%) | 273 | 5 (1.8%) | 181 | 96 | 13.68 | Sig |
| | b) GSS Itobe | 215 | 30 (14%) | 246 | 4 (1.6%) | 397 | 18 | | |
| 6 | a) GSSS Lokoja | 218 | 21 (9.6%) | 215 | 24 (11%) | 282 | 161 | 5.49 | Sig |
| | b) GSS Inye-Ankpa | 130 | 11 (8.4%) | 97 | 12 (12%) | 100 | 25 | | |
| | a) All the six MIP Schools | | | | | 1081 | 773 | 29.50 | Sig |
| | b) All the six non MIP Schools | | | | | 1462 | 186 | | |

Key: N—Number of candidates;
n—Number that pass at least credit level (percentage in parenthesis)
Z—Calculated value of test statistic of a difference between two proportion

The data on Table 4 indicated that the mathematics WASSCE performances of MIP and non-MIP schools are equivalent (in 2009 and 2010) prior to project intervention. This is the main reason that made it possible to compare MIP school results in mathematics WASSCE after project intervention in 2011 with non-MIP school results that are without intervention.

To achieve this, a non-parametric statistics is invoked using the test of significance of a difference between two proportions, and the test statistic is given by:

$$Z = \frac{P_1 - P_2}{\sqrt{\left\{\overline{P}_e \overline{q}_e \left(\frac{N_1 + N_2}{N_1 N_2}\right)\right\}}}$$

Where $P_1$ and $P_2$ are two distinct populations for MIP and non-MIP schools.

$\overline{P}_e = \frac{N_1 P_1 + N_2 P_2}{N_1 + N_2} = \frac{n_1 + n_2}{N_1 + N_2}$ is the variance of the estimated population proportion as suggested by Fisher (Fisher, 1961 & Adetula, 2013c); And $\overline{q}_e = 1 - \overline{P}_e$

To apply these formulas, the illustration of comparing MIP school results in 2011 after the intervention and non-MIP school results without intervention is explicated using S/No 1 on Table 4 above thus:

In the project school (GSSS Icheke Ogane), 31 out of 56 passed, i.e. $P_1 = \frac{31}{56} = 0.55$

In the non-project school (GSS Egume), 97 out of 265 passed, i.e. $P_2 = \frac{97}{265} = 0.36$

$\overline{P}_e = \frac{31+97}{56+265} = \frac{128}{321} = 0.40$, then $\overline{q}_e = 1 - \overline{P}_e = 0.60$

$\therefore \overline{P}_e \overline{q}_e = 0.40 \times 0.60 = 0.24$

Applying the test statistic to get $Z = \frac{0.55 - 0.36}{\sqrt{0.24\left(\frac{56+265}{56\times 265}\right)}} = \frac{0.19}{0.072} = 2.64$

The hypothesis of no significant difference in performance (in mathematics) is rejected beyond the 0.01 level of confidence. Hence, the hypothesis that the two sample proportions arose from the same population is rejected. For others on Table 4, each test statistic was calculated and the decisions were all the same.

**New Direction for MIP**

MIP is introducing electronic classroom mechanism by producing mathematics software (lessons) of different topics into different media like CD, DVD, and memory card (Adetula, 2009). This idea, though at the nascent stage is predicated on the fact that in this 21st Century (electronic and entertaining era), the static media (books and kits) produced by the Centre for teaching and learning may be inadequate; hence the production of mathematics lesson software in different topics becomes a panacea. This is reasonable because in addition to all the advantages the NMC books and kits have, MIP users of this software will be more advantageous to learn at their own pace at their time. Also they can revise the lesson of a topic as many times (slow or fast) as possible and as desired (Adetula, 2012).

With these advantageous positions, many stumbling blocks to mathematics performance will be rooted out. Some of these stumbling blocks to opportunity to learn variables observed and not yet mentioned in the paper include:

- Student over-crowdedness in the classroom. MIP advocates 35 students per class. However MIP school survey revealed 50 to 104 students per class, as indicated on Table1.
- Unreasonable high teaching periods per week for mathematics teachers as a result of fewness in the number of these teachers.
- Even with the few mathematics teachers above, majority of them are not qualified to teach at the grade level they found themselves. Table 1 buttressed this also.
- Most students do not understand Basic English, the language of instruction. Research study supports this, not only at the senior secondary schools but even at the primary/ junior secondary schools that feed these senior secondary (Adetula, 1990).

## Conclusion

That most students fail, dislike and fear mathematics is no longer news. But how this situation can be arrested is the focus of this research project. Therefore in this paper, the author has been able to show that Mathematics Improvement Project (MIP) in Nigeria is an attempt by the Centre to revamp mathematics teaching and learning at the secondary school level. In addition, the Project is designed to present a clear structured and unique teaching methodology that is characterized by active teacher participation and frequent monitoring of students' progress (activities and practices) using NMC germane instructional materials (MIP textbooks series e.g. Basic Mathematics Concepts, Teaching Modules, Students' Workbooks, Models, Games, …) and a "follow through" evaluation mechanism in business-like fashion. This is a "paradigm shift" from the old methodology of expository method to active teaching/ learning, where teachers will able to help students create cognitive maps, link ideas, address misconceptions, and reinforce meaning with teaching models. Such teaching for meaning in mathematics gives students an intellectual foundation to stand upon and a familiar framework to build upon. It also strengthens that framework, as demonstrated in the significance difference in mathematics WASSCE results that favoured MIP schools.

The fundamental conviction of NMC that the mathematics education community (not only in Nigeria but Africa and beyond) must share is that the NMC instructional model can often be used to enrich learning, for it allows students to respond to questions actively and enthusiastically. Note that this visual evidence not only stimulates interest, but help to build mathematical meaning and eventually allow the students to grow and glow mathematically. For all these reasons, it is reasonable to assert that any NMC-MIP teacher will never be involved in presenting a mathematics lesson that is barren, boring, dry, dull, vague, and uninspiring. Such these presentations that allow students to learn mechanical mathematics procedures (mechanistic approaches) for getting answers with little or no understanding to support them are forbidden in MIP teaching episodes.

NMC also noted (Ale & Adetula, 2010), that the most critical element of the education equation is the classroom teacher, and if Nigeria is to have a world –class schools with products of scientific and technological potentials that will achieve Millennium Development Goals, and scientific and technological transformation, then it must have world- class mathematics teaching force. For this reason, the major goal of MIP as clearly stated in the paper is to promote, re-awaken, reactivate, re-energize and re-vigourize the roles of mathematics teachers, as well as the roles of the students to learn mathematics with ease. Consequently, students will excel in mathematics WASSCE with flying colours, as exemplified in Katsina, Kogi, Zamfara and Ondo MIP schools

## References

Adetula, L.O. (1981). Trends in mathematics education in Nigeria. *MAN Bulletin (A publication of Mathematical Association of Nigeria –MAN), 2*(3), 116 – 124.

Adetula, L.O. (1988). Teaching to improve students' problem solving abilities. *Afrika Matematika: A journal of the African Mathematical Union, 2*(1), 139 – 154.

Adetula, L.O. (1989). Teaching and learning mathematics at the Junior Secondary School level. *Nigeria Educational Forum, 12*(1), 197 – 204.

Adetula, L.O. (1990). Language factor: Does it affect children's performance on word problems? *Educational Studies in Mathematics, 21*(4), 351 – 365.

Adetula, L.O. (2002). *Proposal on National Mathematical Centre- Mathematics Improvement Project.* Submitted to National Mathematical Centre. August, 2002.

Adetula, L.O. (2005). National Mathematical Centre – Mathematics Improvement Project (NMC- MIP): A way to enhance student performance in mathematics, *Nigerian Journal of Professional Teachers, 1*(1), 17 – 24.

Adetula, L.O. (2009). The impact of information and communication technology on mathematical sciences education. *Abacus: The Journal of the Mathematical Association of Nigeria, 34*(1), 1 – 6.

Adetula, L.O. (2010). Mathematics: A cognitive tool for national building. In S.O Ale and L.O. Adetula (Eds.), *Reflective and Intellective Position Papers on Mathematics Education Issues* (2nd ed.), (pp. 140 – 149). Abuja: Marvelous Press.

Adetula, L.O. (2012). The unreasonable effectiveness of mathematical reasoning. *The Selected Writings and Articles of L.O. Adetula,* (pp, 29 – 35). Abuja: Marvelous Press.

Adetula, L.O. (2013a). *Framework for Monitoring and Evaluation of Mathematics/ Science Activities in MIP Schools Education Sector Projects.* Abuja: Marvelous Press.

Adetula, L.O. (2013b). Mathematics Improvement Project (MIP): Lesson and self evaluation by ideal MIP mathematics teachers. *NMC – Journal of mathematical sciences education (NMC – JOMSE) 2*(1) 163 – 179.

Adetula, L.O. (2013c). Educational Statistics. In L.O. Adetula (Ed.), *Some Important Topics in Mathematics Education,* (2nd ed.), (pp. 164 – 223). Abuja: Marvelous Press.

Ale, S.O, and Adetula, L.O. (2010). The National Mathematical Centre and the Mathematics Improvement Project in nation building. *NMC Journal of Mathematical Sciences Education 1*(1), 1 – 19.

Deubel, P. *Math Projects: About project-based learning.* Retrieved from http://www.ct4me.net/math_projects.htm

Fafunwa, A.B. (1995). *Speech delivered during the launching of Fafunwa Educational Foundation*. January.

Fisher, R.A. (1961). *The design of experiments.* Edinburgh: Oliver and Boyd.

Larmer, J. and Mergendoller,J. (2010). Seven essentials for project based learning. *Educational Leadership, 68* (1), 34 -37.

Kline, M. (1972). *Mathematical thoughts from ancient to modern times.* NY.: Oxford University Press.

Romberg T.A. (1968). *The development and refinement of prototypic instructional systems.* A paper presented at the symposium of the American Educational Research Association. Chicago, Illinois: February, 1968.

Russell, B. (1968). *The First Volume of Autobiography of Bertrand Russell.* New York: Psychology Press.

Ugoh, S. (1980). *Keynote Address* by the Minister of Science and Technology at the Mathematical Association Annual Conference, Owerri. (August).

Ulm, V. (2011). Teaching mathematics – opening up individual paths to learning. In Series: *Towards new teaching in mathematics, issue 3*. Bayreuth, Germany: SINUS International.

References and citations should be prepared in the APA 6 (http://owl.english.purdue.edu/owl/resource/560/02/ ) format. References have to be cited in article text. See the references examples below.

Angeli, E., Wagner, J., Lawrick, E., Moore, K., Anderson, M., Soderland, L., & Brizee, A. (2010, May 5). General format. Retrieved from http://owl.english.purdue.edu/owl/resource/560/01/

Calfee, R. C., & Valencia, R. R. (1991). *APA guide to preparing manuscripts for journal publication.* Washington, DC: American Psychological Association.

Duncan, G. J., & Brooks-Gunn, J. (Eds.). (1997). *Consequences of growing up poor.* New York, NY: Russell Sage Foundation.

Harlow, H. F. (1983). Fundamentals for preparing psychology journal articles. *Journal of Comparative and Physiological Psychology, 55*, 893-896.

Helfer, M. E., Kempe, R. S., & Krugman, R. D. (1997). *The battered child* (5th ed.). Chicago, IL: University of Chicago Press.

Henry, W. A., III. (1990, April 9). Making the grade in today's schools. *Time*, *135*, 28-31.

Lastname, F. N. (Year). *Title of dissertation*. (Doctoral dissertation). Retrieved from Name of database. (Accession or Order Number)

Lastname, F. N. (Year). *Title of dissertation*. (Unpublished doctoral dissertation). Name of Institution, Location.

O'Neil, J. M., & Egan, J. (1992). Men's and women's gender role journeys: A metaphor for healing, transition, and transformation. In B. R. Wainrib (Ed.), *Gender issues across the life cycle* (pp. 107-123). New York, NY: Springer.

Plath, S. (2000). *The unabridged journals.* K. V. Kukil (Ed.). New York, NY: Anchor.

Schnase, J. L., & Cunnius, E. L. (Eds.). (1995). Proceedings from CSCL '95: *The First International Conference on Computer Support for Collaborative Learning*. Mahwah, NJ: Erlbaum.

Schultz, S. (2005, December 28). Calls made to strengthen state energy policies. *The Country Today*, pp. 1A, 2A.

Scruton, R. (1996). The eclipse of listening. *The New Criterion, 15*(30), 5-13.


International Journal of Contemporary Educational Research (IJCER)

www.ijcer.net

# Using Problem-Based Learning to Improve College Students' Mathematical Argumentation Skills

**Bambang Aryan Soekisno[1], Yaya S. Kusumah[2], Jozua Sabandar[2], Darhim[2]**

[1]STKIP Siliwangi
[2]Department of Mathematics Education , Indonesia University of Education

**To cite this article:**

Soekisno, B.A., Kusumah, Y.S., Sabandar, J., & Darhim. (2015). Using problem-based learning to improve college students' mathematical argumentation skills. *International Journal of Contemporary Educational Research*, *2*(2), 118-129.

IJCER International Journal of Contemporary Educational Research
Volume 2, Number 2, July 2015, Page 118-129 ISSN: 2148-3868

# Using Problem-Based Learning to Improve College Students' Mathematical Argumentation Skills

**Bambang Aryan Soekisno[1*], Yaya S. Kusumah[2], Jozua Sabandar[2], Darhim[2]**
[1]STKIP Siliwangi
[2]Department of Mathematics Education, Indonesia University of Education

## Abstract

Problem-based learning are very common in mathematics teaching and learning in college level where students are exposed to problems and required to solve them. However, in general, when students are making efforts to solve the problems, they usually implement direct problem solving strategy without involving argumentation as one of the most important component of the problem solving process. Through argumentation, a reasoning process will take place and will involve thinking about what data gathered, what theorem support the process, how rebuttal is given, and what final claim to confirm. A student is able to comprehend a certain problem meaningfully when he/she is able to state the reasons, to elaborate the data, to express assurance, and even to make a claim of a problem in correct ways. Therefore, in order to see if the students have already acquired the skills to express mathematics problem meaningfully, they can be asked to express ideas orally and rewrite the ideas in mathematical arguments. Thus, this research was intended to find out if the students have improved their mathematical argumentation in Calculus 1 course. A learning process that facilitates the students to develop arguments is needed in order to improve their mathematical argumentation skills. Therefore, this research took students at the Dept. of Mathematics Education in UHAMKA University as the subjects in which four classes of them were taken as the samples using purposive random sampling technique. The four classes were then divided into two groups: two classes as control groups and two other classes as experimental groups. In this research, the teaching and learning process in the experimental groups used Problem-based Learning (PBM); meanwhile, conventional learning process (KS) was implemented in the control groups. There were 141 students involved in this research. And the instrument used in this research was a test intended to measure mathematical argumentation skills. The data was analyzed using *t*-test and ANOVA through one and two lines. Based on the analysis, this research found that there was significant differences in mathematical argumentation between PAM Groups (upper, middle, and lower) on Problem-based Learning. There was differences in the improvement of PAM for upper and middle groups. Significantly, the improvement of students' mathematical argumentation skills in PAM groups using Problem-based Learning is better than those using conventional learning. There was also significant improvement on students' mathematical argumentation skills in each of PAM groups between those using Problem-based Learning and Conventional Learning. Collectively, both factors of PM groups and learning approach have given significant effects on students' mathematical argumentation skills.

**Key words:** Mathematical argumentation skills, Problem-based learning

## Introduction

The ability to express ideas supported by data and based on sufficient theories of a mathematical problem, both oral and written, is an important part of mathematical abilities that students in college level need to have. Ideas that are supported by sufficient data and based on appropriate theories would provide students with sufficient understanding about mathematical concepts. Sufficient reasons support explanations about something to be considered as right or wrong. In addition, reasons would also support an interpretation of many concepts. Changes would happen when one changes their understanding on certain concepts and conceptual framework they use. It also may happen when one redesign the framework to accommodate new perspectives.

The ability to express mathematical problems can be seen when students convey their ideas orally and rewrite their ideas into mathematical arguments. Ideas about optimizing the ability to make arguments in mathematics

* Corresponding Author: *Bambang Aryan Soekisno, bambang_aryan@yahoo.com*

such as expressing reasons, data, theories, writing, and developing discourses have become one of the alternatives of finding answers on certain problems. That is why, it is not impossible that a doer and a user of mathematics look for forms, models, and trick during working on mathematics problems.

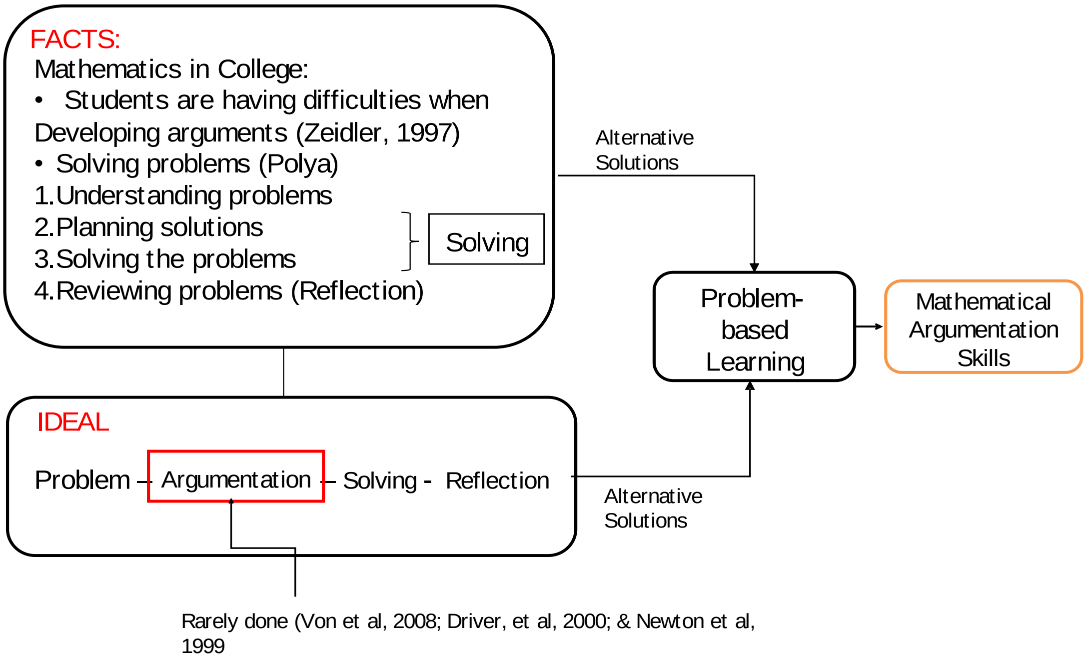
The process of searching for solutions of problems is surely not a simple reasoning process. It needs higher level of reasoning. The ability to collect information and data, to express arguments, to determine supporting theories, to decide the pathway of problem solving process, have become part of reasoning process that make students possible to solve the problems.

The ability to make an argumentation is an important part of mathematics learning. Therefore, students need to practice it as often as possible. They need to have argument skill to make them able to solve problem critically. Argumentation are the essence of scientific reasoning (Cross, 2007).

The ability is also a foundation of logical and critical thinking. According to Ennis (1981), critical thinking is an ability to express reasons based on what one beliefs. The argumentation ability involves skills to express a reasons critically supported by data and theories of mathematics problems (Logic).

Argumentation as a foundation of critical thinking and logic is still viewed as something difficult by students (Zeidler, 1997). Von et al. (2008), Driver, et al. (2000) and Newton, et al. (1999) said that the difficulties the students encounter when developing argumentation is caused by lack of pedagogical skills of lecturers in developing argumentation in classroom. Students often find difficulties when developing critical arguments due to poor learning process in classroom. However, in classroom, college students are often encountered with problems that they need to find the solution through argumentation process.

Before arriving to the solving problem phase, students need to go through a series of steps including reasoning process, data gathering, identifying theorem, rebuttals, and make some claims. Later, the problems can be solved with hopes that the process of solving problems are in the right directions.




Picture 1. Logical Framework for Improving Students' Mathematical Argumentation Skills

Argumentation skills are critical and logical thinking about the relationship between concepts and situations. The skills are very useful to explain about facts, procedures, concepts, and methods of solving problems that are interrelated one another. It is expected that the higher the students' mathematical argumentation skills, the better their ability to express reasons for certain solutions or answers.

Learning process needs to change in order to prepare students to face new situations. Students' skills in asking questions, searching and finding for appropriate resources to answers the questions, and communicating effectively the solutions that they have from others. Problem-based learning is an approach in mathematics education that helps students develop critical and logical thinking and other skills needed in order to be able to communicate successfully.

Arguments are rational ways that one take to answer questions, to face issues, and to rebut and solve problems. An argument consists of a claim (Solution) that is supported by many principles (Assurance), proofs, and rebuttals that contradict with other potential arguments. Developing an argument in learning process can potentially solve problems.

Mathematical argumentation are important when one learns about how to solve many kind of big problems. It is also a powerful method that can be used to measure the ability in solving problems. The arguments are also useful to solve both poor-structured and well-structured problems (Jonassen, 2010). Moreover, Nussbaum & Sinatra (2003) said that students need to show that they are able to improve their reasoning over problems when they find solutions; or when developing an arguments in order to find appropriate answers scientifically.

There are many factors inhibit learning process such as students' lower reasoning level including argumentation. The most common weakness that the students have in making argumentation is in counter-argument. When a student is asked to make arguments to support and challenge something, he/she usually elaborate many reasons to support his/her positions (Stein & Bernas, 1999). Qualification and rebuttal are rarely used in analyzing arguments in mathematics education; and they are very useful in analyzing arguments made by students (Inglis et al, 2007).

There are many efforts have been done in many developed countries intended to find the causes and solutions to problem of lack mathematical argumentation skills of students. The efforts were done using many kinds of educational theories, learning models and approaches which resulted in improvement of knowledge about argument. Experts such as Conner (2008) provides a description about correlation between argument and proofs in geometry class. Meanwhile, Halpren (2003) identifies a process intended to analyze students' argument elaboration that involves reading and evaluating the argument based on the strength of correlation between premises, conclusions, assumptions, and counter-arguments.

Through problem based learning, students are expected to think critically, to analyze complex problems and real-life problems, to work cooperatively in small groups, to be skillful in making effective and accurate communication both verbal and non-verbal in order to show that they have mathematical arguments. Students' mathematical arguments will improve if they are involved in problem based learning, especially unstructured problems and interpretations over alternative solutions that needs argumentation. Students are required to remember information with minimum reasons in order to involve in making argumentation. Problem based learning environment usually exposes students to claims and alternative solutions that need to handle by them through argumentation.

From the perspective of pedagogy, the objective of learning for adults is to provide opportunities to students to doing math. Nowadays, learning focuses on the use of surrounding environment as the source such as in problem based learning. Problem based learning is a classroom activity intended to organize learning about problem solving. It is also about providing opportunities to students to express their arguments and mathematical ideas, and to communicate with their peers through interaction of components in classroom. This is in accordance with NCTM (2000) which says that developing learning environment that challenge and support important learning component is critical.

A learning process that provide many opportunities to students to develop arguments called as problem based learning. The learning begins with exposing students to many contextual problems, current problems, and other hot problems. The problem exposed in the initial stage is usually word-problem along with direction to develop counter-argument. Counter-argument is defined as attribute of a good argumentation (Andriessen et al, 2003; Voss et al, 1991) and a standard for measuring an argument (Kuhn, 1991).

Problem based learning mostly provides students with opportunities to make argumentation. This learning is expected to guide students to reach the objectives of the learning namely the ability to make mathematical argument in a discussing a mathematical problems. This research took students of Mathematics education, especially those who take Calculus I. The reason why choosing this topic was that Calculus I provides many mathematical problems taken from real-life situation.

## Method

### Participants

Subjects in this research were students of mathematics education of UHAMKA University. They were chosen for their enrollment in Calculus I course – offered for freshmen. Other considerations when choosing these group of students were that they offered heterogeneity of academic skills, their level of reasoning, and their independent in learning. All these factors surely influenced the implementation of problem based learning.

The samples of this research were students at Mathematics education department of UHAMKA University who enrolled in Calculus I course. The samples were drawn using purposive random sampling for its excellence in providing randomized and more variant samples. All of the students were then divided into two groups: Experimental group and control group. In order to assure the randomized result of the samples, this research employed drawing techniques.

### Research Procedures

This research was a quasi-experimental study about the using of problem based learning. There were two group of student involved in this study in which their mathematical argument skill was measured in order to see if the problem based learning had successfully improve the students' mathematical argument skills. The first group called experimental group and they received treatment using problem based learning; and the second group called control group and they received treatment using conventional teaching method. In both groups, the students were then categorized into the following mathematical argument skills level as follows: higher, middle, and lower.

There were two stages in this research: identifying and developing learning components, and conducting the research that covered the whole planned process.

### Instruments and Sources of Data

This research employed several instrument to collect data. The instruments were as follows: (a) a test of students' initial skill level of mathematics; (b) a test of students' initial mathematical argument skill; (c) observation sheet; and (d) interview protocol intended to collect students' perspective towards problem based learning. Students' mathematical argument skill were measured through written test that consists of the following aspects: identification of assumption, identification of relevance and irrelevance data, analysis of argument, answer question with reasons (clarification), and give reasons on certain conclusion. Meanwhile, the test of mathematical argument skills consisted of eight items. These instruments were given to the students before and after the learning process.

### Data Analysis

This research analyzed the data using two approaches: quantitative and qualitative. Quantitative analysis consisted of descriptive and inferential statistics. The first step was descriptive statistics analysis included finding out the mean and standard deviation; then a diagram graphic was used to see the general description. In order to find out if there was improvement on students' mathematical arguments on both groups of samples, this research conducted an analysis on the results of pretest and posttest. Moreover, the data analysis was conducted by using the formula of average normalized gain (Hake, 2007).

The second step of data analysis was inferential statistics analysis that was carried out to test the hypotheses. This step was begun by conducting normality and homogeneity of variant both partly and wholly. Later, in order to find out if there were differences in the groups, to check if there was an interaction between free variable and controlled variable related to hypothesis, this research employed ANOVA with one line using SPSS-19.00 software with 95% level of trusted.

## Research Findings

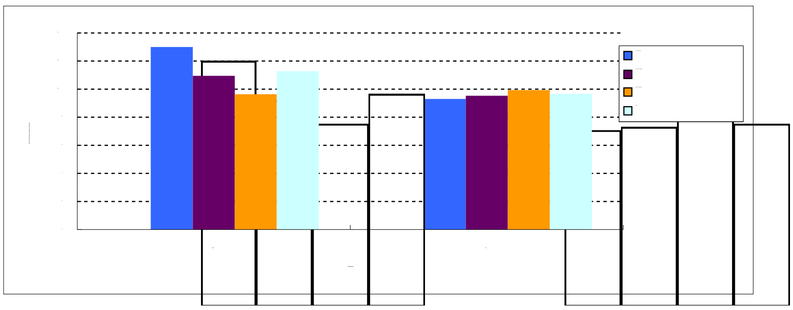
This parts highlights the research findings. The analysis conducted in this research was intended to find out if there was interaction between the treatments, the students' initial level of skills of mathematical argumentation (PAM), and their mathematical argumentation skills.

**Table 1. Sample Distribution**

| PAM Groups | Experimental Group | Control Group | Total |
|---|---|---|---|
| Upper | 18 | 10 | **28** |
| Middle | 40 | 32 | **72** |
| Lower | 11 | 30 | **41** |
| Total | **69** | **72** | **141** |

Quantitative data was gathered from pretest intended to find out students' initial mathematical skills and their mathematical argumentation. There were 141 students taken the test in which 69 students in experimental group and 72 students in control group.

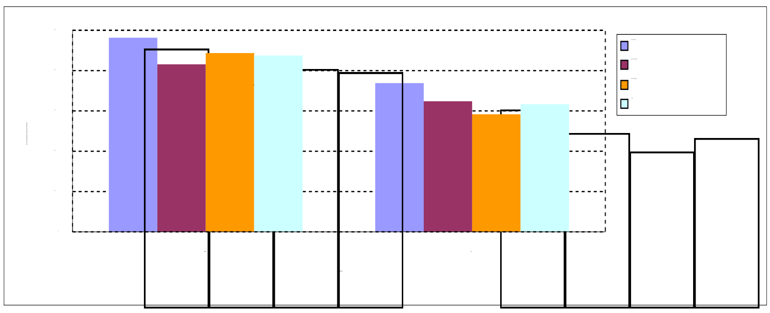
Picture 2 and 3 displays the scores of students' mathematical argumentation before and after the learning process. Meanwhile, the scores of students' mathematical argumentation for each PAM are summarized in Picture 2.



Notes: Ideal maximum score is 70

**Picture 2. Average of Students' Initial Mathematical Argumentation based on PAM Groups before Treatment.**

The following Picture 3 displays the scores of students' mathematical argumentation based on PAM Groups after the treatment.



Note: Ideal Maximum Score is 70

**Picture 3. Average of Students' Final Mathematical Argumentation based on PAM Group after the Learning Process**

The following Table 2 summarizes the improvement of scores of students' mathematical argumentation before and after the learning process.

**Table 2. The Description of n-Gain of Mathematical Argumentation Skills based on Learning and PAM Groups**

| PAM | Learning Process | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Problem based Learning | | | | | Conventional Learning | | | | |
| | Minimum Scores | Maximum Scores | Mean | SD | n | Minimum Scores | Maximum Scores | Mean | SD | n |
| Upper | 0,456 | 0,763 | **0,618** | 0,091 | 18 | 0,379 | 0,574 | **0,455** | 0,066 | 10 |
| Middle | 0,317 | 0,737 | **0,521** | 0,102 | 40 | 0,121 | 0,571 | **0,377** | 0,114 | 32 |
| Lower | 0,452 | 0,770 | **0,575** | 0,105 | 11 | 0,119 | 0,525 | **0,320** | 0,110 | 30 |
| Total | 0,317 | 0,770 | **0,555** | 0,107 | 69 | 0,119 | 0,574 | **0,364** | 0,115 | 72 |

The Table 2 above indicates that there is improvement in students' mathematical argumentation in which students in experimental group (Problem based learning) had better scores than those in control group (Conventional). The average scores (Mean) of n-gain of mathematical argumentation skill in experimental group is 0.555. it means that the mean scores of n-gain of mathematical argumentation of experimental group is higher than those in control group with only 0.36. In addition, the mean scores of improved mathematical argumentation in experimental group for all PAM groups (Upper, Middle, and Lower) is also higher than those in control group.

This general description indicates that there is a difference on mathematical argumentation skill between students in experimental and control group. Then, a statistical test was carried out to see if the difference is significant.

In order to see if there is difference on the mean scores of improved mathematical argumentation between experimental and control group, this research conducted *t*-test as displayed in the following Table 3.

**Table 3. t-Test on n-Gain Average Scores of Mathematical Argumentation Skill gained by Experimental and Control Groups**

| Argumentative Skills | t-test | | | H0 |
|---|---|---|---|---|
| | T | df | Sig. | Rejected |
| | 10.174 | 139 | 0.000 | |

Before testing the hypotheses, all requirements must be met. The tested hypothesis was: H0: There is no significant difference on mean scores of students' mathematical argumentation skills gained by experimental and control group. Ha: The mean scores of students' mathematical argumentation skills gained by experimental group are higher than those in control group. The testing criteria said: if the probability scores (Sig.) is higher than 0.05, then the null hypothesis is accepted.

The following Table 3 shows that the null hypothesis is rejected. Therefore, it can be concluded that the improved mathematical argumentation skills gained by the students in experimental group is significantly higher than those in control group.

**Table 4. ANOVA Difference of Mean Scores (Average) of Improved Mathematical Argumentation based on PAM Groups and Problem base Learning.**

| Difference resources | Sum of Squares | df | Mean Square | F | Sig. | $H_0$ |
|---|---|---|---|---|---|---|
| Between Groups | 0.124 | 2 | 0.062 | 6.222 | 0.003 | |
| Within Groups | 0.656 | 66 | 0.010 | | | Rejected |
| Total | 0.779 | 68 | | | | |

The n-gain data scores of students' mathematical argumentation both total scores and PAM based scores in experimental group are normally distributed and homogeneously varied. One line ANOVA was employed to find out if there was a difference in the improvement of students' mathematical argumentation based on PAM group and in problem based learning classroom. The summary of the ANOVA analysis can be seen in the Table 4.

Table 4 shows that probability scores (Sig.) is 0.003 which means that the average improvement of students' mathematical argumentation skills among the PAM groups in problem based learning classroom is significantly difference to one another. Then, this research conducted the Scheffe Test to find out which improvement that significantly difference in the students' mathematical argumentation skills. The summary of the test can be seen in the following Table 5.

**Table 5. Scheffe Test on the Average Scores of Improved Mathematical Argumentation based on PAM Groups in Problem Based Learning Classroom**

| (I) PAM Groups | (J) PAM Groups | Mean Difference (I-J) | Std. Error | Sig. | 95% Confidence Interval | | $H_0$ |
|---|---|---|---|---|---|---|---|
| | | | | | Lower Bound | Upper Bound | |
| Upper | Middle | 0.098* | 0.028 | 0.004 | 0.027 | 0.169 | Rejected |
| | Lower | 0.043 | 0.038 | 0.526 | -0.052 | 0.139 | Accepted |
| Middle | Upper | -0.098* | 0.028 | 0.004 | -0.169 | -0.027 | Rejected |
| | Lower | -0.054 | 0.034 | 0.286 | -0.139 | 0.031 | Accepted |
| Lower | Upper | -0.043 | 0.038 | 0.526 | -0.139 | 0.052 | Accepted |
| | Middle | 0.054 | 0.034 | 0.286 | -0.031 | 0.139 | Accepted |

*. The mean difference is significant at the 0.05 level.

The tested Hypotheses were as follows:

$H_0$: There is no significant difference on mean scores of students' mathematical argumentation skills gained by experimental and control group..

$H_a$: The mean scores of students' mathematical argumentation skills gained by experimental group are higher than those in control group.

The testing criteria said that if the probability scores (Sig.) is higher than $\alpha = 0.05$, then the hypothesis is accepted.

Based on the computation result in Table 5, it can be seen that there were probability scores (Sig.) for each pairs of PAM group: upper and middle, upper and lower, and middle and lower. The probability scores for the pair of Upper and Middle PAM group is 0.05 which means that the null hypotheses is rejected. Therefore, it can be said that there is significant improvement on mathematical argumentation skill of the Upper and Middle PAM Group. Meanwhile, the similar findings also happened in the pairs of Upper and Lower PAM group in which the probability scores is 0.05 and the null hypothesis is accepted. Therefore, it can be concluded that there is no significant different in students' mathematical argumentation skill between Upper, Middle, and Lower PAM groups.

This research also found that the overall data of students' mathematical argumentation skills were taken from each PAM Groups and gathered through normal distribution with homogeneous variants. This research then conducted a two lines ANOVA test in order to find out if there was interaction between learning process and PAM groups in mathematical argumentation skills.

According to the computation results of ANOVA as indicated in Table 6, it was found that the F Scores for PAM groups are 7.402 with probability values (Sig.) in 0.001. This figure means that PAM Groups have given significant effect on the improvement of students' mathematical argument skills and so is the factor of learning approach. It can be seen from the F scores of learning approach with 87.426 and probability value (Sig.) in 0.000 which means that this significant values are less than 0.05 thus making the null hypothesis is rejected.

According to the computation results displayed in Table 6, the F Scores for interaction between PAM Groups and learning approach is 3.190 with probability values (Sig.) in 0.444. This score is less than 0.05 thus the null hypothesis is rejected. Therefore, it can be concluded that the PAM Group factors and learning approach have jointly given significant impact on students' mathematical argumentation skills.

The following Table 6 summarizes the result of the two lines ANOVA test.

**Table 6. The Two Lines ANOVA between PAM Groups and Students' Mathematical Argumentation Skills**

| Source | Type III Sum of Squares | df | Mean Square | F | Sig. | $H_0$ |
|---|---|---|---|---|---|---|
| Corrected Model | 1.545[a] | 5 | .309 | 28.777 | .000 | |
| Intercept | 24.501 | 1 | 24.501 | 2281.022 | .000 | |
| PAM | .159 | 2 | .080 | 7.402 | .001 | Tolak |
| Learning | .939 | 1 | .939 | 87.426 | .000 | Tolak |
| PAM * Learning | .069 | 2 | .034 | 3.190 | .044 | Tolak |
| Error | 1.450 | 135 | .011 | | | |
| Total | 32.583 | 141 | | | | |
| Corrected Total | 2.996 | 140 | | | | |

a. R Squared = .516 (Adjusted R Squared = .498)

This research continued to conduct Scheffe test to find out which PAM group interacts with learning approach in improving students' mathematical argumentation skills. The test results are summarized in Table 7 below. The table shows that the improvement of students' mathematical argumentation skills in Upper PAM Group is better than in Middle and Lower PAM Groups. Meanwhile, the improvement of students' mathematical argumentation in Middle PAM group is better than in Lower PAM group. It indicates that problem based learning (PAM) has played significant roles in improving students' mathematical argumentation skill. In addition, the difference of to which degree the improvement occur between problem based learning classroom and conventional classroom is significant in which the probability values (Sig.) is less than 0.05. It means that there was interaction between learning factors (Problem based and conventional) and PAM Groups (Upper and Middle) in improving students' mathematical argumentation skills.

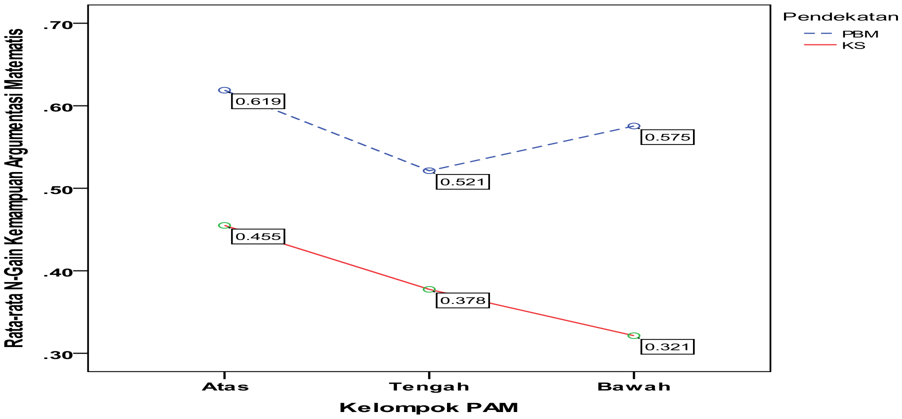
**Table 7. The Comparison of Improvement of Students' Mathematical Argumentation in both Problem based Classroom and Conventional Classroom**

| (I) PAM Groups | (J) PAM Groups | Mean Difference (I-J) | Std. Error | Sig. | 95% Confidence Interval Lower Bound | 95% Confidence Interval Upper Bound | $H_0$ |
|---|---|---|---|---|---|---|---|
| Upper | Middle | 0.103* | 0.023 | 0.000 | 0.046 | 0.160 | Rejected |
| | Lower | 0.171* | 0.025 | 0.000 | 0.108 | 0.234 | Rejected |
| Middle | Upper | -0.103* | 0.023 | 0.000 | -0.160 | -0.046 | Rejected |
| | Lower | 0.068* | 0.020 | 0.005 | 0.018 | 0.118 | Rejected |
| Lower | Upper | -0.171* | 0.025 | 0.000 | -0.234 | -0.108 | Rejected |
| | Middle | -0.068* | 0.020 | 0.005 | -0.118 | -0.018 | Rejected |

*. The mean difference is significant at the 0.05 level.

The research then conducted data analysis using graphic intended to find out the interaction between learning process and PAM Group in improving students' mathematical argumentation skills. The graphic had helped this research in identifying the interaction as it can be seen in Picture 4.

Picture 4 above illustrates the interaction between learning factors (Problem based Learning and Conventional Learning) and the PAM groups in improving students' mathematical argumentation skills. This interaction happens due to the gap in the improvement of students' mathematical argumentation between problem based classroom and conventional classroom. Therefore, it can be concluded that there were interactions between learning factors (Problem based learning and conventional learning) and PAM Groups (Upper, Middle, and Lower) in improving students' mathematical argumentation.




**Picture 4. The Interaction between Learning Approach and PAM Groups in Improving Students' Mathematical Argumentation Skills**

## Discussions

This research has successfully found that there was significant improvement in students' mathematical argument skills in which students in experimental group have better improvement than those in control group. Table 2 highlights the average of improvement of the argumentative skills where students in experimental group gained the scores of 0.555 (moderate improvement) and students in control groups gained the scores of 0.364 (Light improvement). This improvement is supported by the statistical test results which indicate that mathematical argumentation skills of students in problem based learning classroom is better than those in conventional learning classroom.

The improvement of students' mathematical argumentation skills happens due to the implementation of problem based learning that exposes students with argumentative skills. The more often the students practice their argumentation skills, the more skillful they in making argumentations. Osborne (2005) said that making an argument as a long process that needs experience and significant practice repetitively. In addition, the improvement of argumentation skills also happens due to the opportunities that problem based learning provide to the students have to understand basic knowledge, facts, and application and the ability to communicate effectively and accurately both oral and written, and to work together in small and big groups (Duch et al 2001).

In average, the improvement scores of the students' mathematical argumentation skills are 0.618 for Upper PAM group, 0.521 for Middle PAM Group, and 0.575 for Lower PAM group to which all of them considered to be moderate improvement category. The improvement of the mathematical argumentation skills gained by students in problem based learning class for Upper PM group is better than those in Middle and Lower PAM group. But, students in Middle PAM group did not get better improvement than those in Lower PAM group. In average, students in Middle PAM group get lowest improvement among all PAM groups.

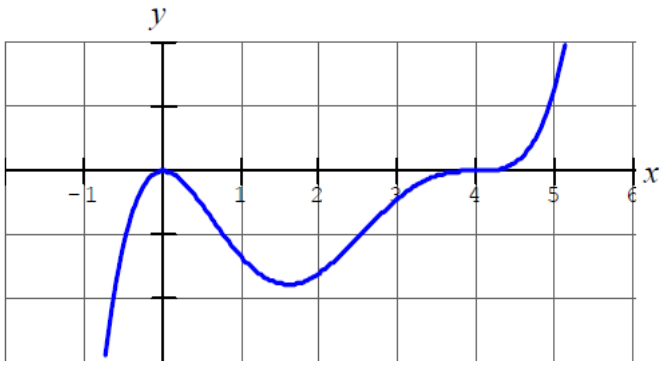
Meanwhile, in the conventional classroom, the average improvement of mathematical argumentation skills for Upper PAM was 0.455, for Middle PAM was 0.377, and 0.320 for Lower PAM group, and this improvement can be categorized into light improvement. In addition, students in Upper PAM get the highest improvement among three PAM groups.

In the problem based classroom, there was significant difference of improvement gained by student in the three groups (Upper, Middle, and Lower). The most significant difference occur between Upper and Middle group. This research also found that improvement gained by students in Middle group of Problem based classroom is higher than the upper group in conventional classroom. This is based on the data analysis results which indicated that the improvement of Middle group in Problem based classroom was 0.521; meanwhile in the Upper group of conventional classroom, the scores were 0.455. And the improvement in Middle PAM of problem based classroom and Upper PAM in conventional classroom can be categorized into light/moderate improvement.

Based on the argumentation level, the Toulmin model of students' responses on the question Number 7a is in the level of 5.

**Question Number 7a.**

Look at the following $f(x)$ function:

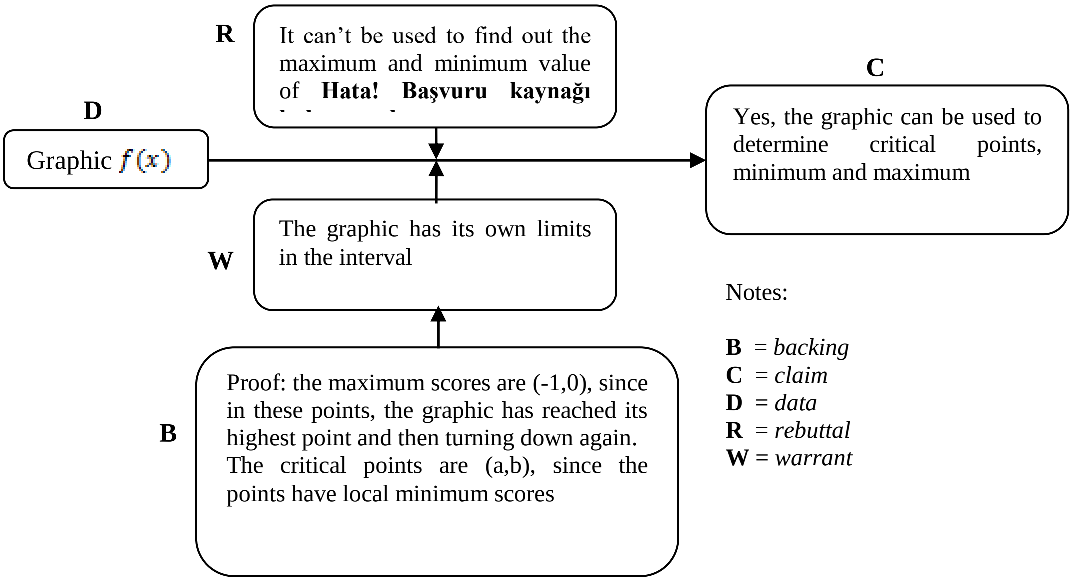



a) If the graphic displayed is graphic of $f(x)$. Can the graphic be used to determine critical point, local maximum and minimum, and absolute minimum of $f(x)$? Provide your reasons!

The students are able to understand the concept of minimum and maximum of a curve. Students' responses indicate that they have already possessed the pathway of thinking since it consisted of data, claim, warrant, backing, and rebuttal.

| | | |
|---|---|---|
| Data | : | *Graphic based questions* |
| Claim | : | *Yes, the graphic can be used to determine critical point, maximum, and minimum.* |
| Rebuttal | : | *But it can't be used to find out the absolute value of minimum and maximum of f(x).* |
| Warrant | : | *The graphic has its own limit in each interval..* |
| Backing | : | *Proof: the maximum score is (-1,0), since in the point, the graphic reach the highest point before turnig down again. The critical points are (a,b), since in these points, the graphic has its local minimum scores.* |

The following picture visualizes the students' responses:




Notes:

**B** = *backing*
**C** = *claim*
**D** = *data*
**R** = *rebuttal*
**W** = *warrant*

The results of data analysis have informed that there was interaction between PAM groups (Upper, Middle, and Lower) and learning factors in improving students' mathematical argument skills. Both PAM groups and learning factors have given impact significantly on the improvement of mathematical argumentation of the students. The findings of the research indicate that there is a huge possibility to implement problem based learning in improving students' mathematical argumentation skills.

In addition, this research also found that there was difference improvement of argumentation skills between students in the three PAM groups. The most significant different is between Upper and Lower PAM group. It indicates that problem based learning is appropriate to implement in Upper PAM group. As it can be seen from its characteristics, through learning in a problem based classroom, students in Upper PAM group can optimize their communication skills to be more effective and accurate both verbal and non-verbal. Students who have higher initial mathematical skills (Upper group) have had experienced activities intended to solve mathematical problems and in real-life situation. Thus, they are able to make a claim through logical analysis and be able to provide supporting logical evidences.

## Conclusion

In conclusion, the improvement of students' mathematical argumentation skills in problem based learning classroom is better than those in conventional learning classroom. The average improvement gained by students in all groups (Upper, Middle, and Lower) in problem based learning are significantly different to one another. Moreover, by using Scheffe Test, this research identified that there was significant difference in the improvement between Upper group and Middle group. The factors of PAM group have significantly given impact on the improvement of students' mathematical argumentation skills. There are also interactions between learning factors and PAM group. Looking at the scores, problem based learning found to be more appropriate to students in Upper and Lower group in improving students' mathematical argumentation skills.

## References

Andriessen, J., Baker, M., & Suthers, D. (2003). Argumentation, computer support, and the educational context of confronting cognitions. In J. Andriessen, M. Baker & D. Suthers (Eds.), *Arguing to learn: Confronting cognitions in computer-supported collaborative learning environments* (pp. 1-25). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Conner, A. (2008). Expanded Toulmin diagrams: A tool for investigating complex activity in classrooms. In O. Figueras, J. L. Cortina, S. Alatorre, T. Rojano, & A. Sepulveda (Eds.), *Proceedings of the Joint Meeting of the International Group for the Psychology of Mathematics Education* 32 *and the North American Chapter of the International Group for the Psychology of Mathematics Education* XXX. Vol. 2. (pp. 361-368). Morelia, Mexico: Cinvestav-UMSNH.

Cross, D., (2007) *Creating Optimal Mathematics Learning Environments: Combining Argumentation and Writing to Enhance Achivement*. Disertasi University of Georgia: Tidak diterbitkan.

Driver, R., Newton, P., and Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education*, 84(3), 287–312.

Duch, B.J., Groh, S.E., dan Allen, D.E. (2001). Why Problem-Based Learning: A Case Study of Institutional Change in Undergraduate Education. Dalam B.J. Duch, S.E. Groh, dan D.E. Allen (Eds): *The Power of Problem-Based Learning*. Virginia, Amerika: Stylus Publishing.

Ennis, R.H. (1981). *Critical Thinking*. United States of America: Prentice-Hall, Inc.

Hake, R.R. (2007). *Design-Based Research in Physics Education Research*: A Review, in A.E. Kelly, R.A. Lesh, & J.Y. Baek, eds. (in press), Handbook of Design Research Methods in Mathematics, Science, and Technology Education.

Halpern, D. F. (2003). *Thought and Knowledge: An Introduction to Critical Thinking* (4th ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Inch, E.S., Warnick, B., & Endres, D. (2006). *Critical Thinking and Communication: The Use of Reason in Argument*. Boston: Pearson Education Inc.

Inglis, M., Mejia-Ramos, J.P., & Simpson, A. (2007). Modelling Mathematical Argumentation: The Importance of Qualification. *Educational Studies in Mathematics.*

Jonassen, D.H. (2010). *Learning to Solve Problem: An instructional guide design*. San Fransisco: Pfeiffer

Kuhn, D. (1991).*The skills of argument*. Cambridge University Press.

National Council of Teachers of Mathematics (2000). *Principles and Standards for School Mathematics*. [Online]. Tersedia:http:// www.nctm.org/ standars/overview.htm [25 Januari 2011]

Newton, P., Driver, R., & Osborne, J. (1999). The Place of Argumentation in The Pedagogy of School Science. *International Journal of Science Education*, 21(5), 553–576.

Nussbaum, E. M., & Sinatra, G. M. (2003). Argument and Conceptual Engagement. *Contemporary Educational Psychology, 28,* 384-395.

Osborne, J. (2005). The Role of argument in Science Education. K. Boesma, M. Goedhart, O. De Jong, & H. Eijkelhof [Eds]. *Research and Quality of Science Education.* Dordrecht, Nederlands: Spinger.

Stein, N., & Bernas, R. (1999). The Early Emergence of Argumentative Knowledge and Skill. In J. Andriessen & P. Coirier (Eds), *Foundations of Argumentative Text Processing* (pp. 97-116). Amsterdam: Amsterdam University Press.

Toulmin, S.E. (2003). *The Uses of Argument*. New York: Cambridge University Press

Von Aufschnaiter, C., Erduran, S., Osborne, J. & Simon, S. (2008). Arguing to Learn and Learning to Argue: Case Studies of How Students' Argumentation Relates to Their Scientific Knowledge. *Journal of Research in Science Teaching, 45(1),* 101-131.

Voss, J.F., Perkins, D.N., & Segal, J.W. (1991). *Informal Reasoning and Education*. Hillsdale , NJ: Erlbaum.

Zeidler, D. L. (1997). *The Central Role of Fallacious Thinking in Science Education*. Science Education, 81, 483–496