12-2012

Which Study Method Works Best? A Comparison of SOAR and SQ3R for Text Learning

Sarah C. Kasson
University of Nebraska-Lincoln, sarahcr87@gmail.com

Follow this and additional works at: http://digitalcommons.unl.edu/cehsdiss

Part of the Educational Psychology Commons, and the Psychology Commons

http://digitalcommons.unl.edu/cehsdiss/165

This Article is brought to you for free and open access by the Education and Human Sciences, College of (CEHS) at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Public Access Theses and Dissertations from the College of Education and Human Sciences by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
WHICH STUDY METHOD WORKS BEST?
A COMPARISON OF SOAR AND SQ3R FOR TEXT LEARNING

by
Sarah C. Kasson

A THESIS

Presented to the Faculty of
The Graduate College at the University of Nebraska
In Partial Fulfillment of Requirements
For the Degree of Master of Arts

Major: Educational Psychology

Under the Supervision of Professor Kenneth A. Kiewra

Lincoln, Nebraska

December, 2012
WHICH STUDY METHOD WORKS BEST?

A COMPARISON OF SOAR AND SQ3R FOR TEXT LEARNING

Sarah C. Kasson, M.A.

University of Nebraska. 2012

Adviser: Kenneth A. Kiewra

One hundred thirty-eight college students participated in a study comparing the SOAR (Select, Organize, Association, Regulate) and SQ3R (Survey, Question, Read, Recite, Review) study systems to each other and to students’ preferred study methods. Though both systems have been researched independently, just one other study has compared these study systems to each other. College students were assigned randomly to one cell of a 2x2 factorial design (method: SOAR or SQ3R; material: supplement or no supplement) or to a preferred-study-method control group. Groups were trained in their respective system (SOAR, SQ3R, or control) and then given materials about educational measurement to study. During the study period, half of SOAR and SQ3R trained students were given a study supplement that matched their training method. The other halves and the control group were not given supplements. Following the study period, participants were tested on the educational measurement material with respect to fact, relationship, concept, and skill learning. Finally, all participants completed an attitudinal survey regarding their experiences. Achievement results showed no main or interactive effects for method or material and no mean differences among the three study methods (SOAR, SQ3R, and control). Results from the attitude survey revealed that students trained in SQ3R felt more prepared to study than those who received SOAR or control training.
Dedications

For Chuck, who puts up with my particular brand of crazy and loves me anyway.

For my parents, who have always believed in me and have been there with never-ending patience and support.

For Megan and my margarita girls, who listen to me complain without complaint.

For my friends (A, E, K, and M) without whom I never would have made it through.

For Dr. Kiewra, who was onboard when I needed him most.

This would not have been possible with your love, feedback, and support.

Special thanks to Mike Zweifel at the NEAR Center and to Linlin Luo, my statistical gurus.

Thank you all from the bottom of my heart and the tips of my toes.

-Sarah
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>List of Tables</th>
<th>v</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Chapter</strong></td>
<td></td>
</tr>
<tr>
<td>I. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>II. Literature Review</td>
<td>12</td>
</tr>
<tr>
<td>III. Methods</td>
<td>31</td>
</tr>
<tr>
<td>IV. Results</td>
<td>37</td>
</tr>
<tr>
<td>V. Discussion</td>
<td>47</td>
</tr>
<tr>
<td>VI. References</td>
<td>53</td>
</tr>
<tr>
<td>VII. Appendix A: Demographic Information</td>
<td>61</td>
</tr>
<tr>
<td>VIII. Appendix B: Educational Measurement Text</td>
<td>62</td>
</tr>
<tr>
<td>IX. Appendix C: SQ3R Ideal Materials</td>
<td>69</td>
</tr>
<tr>
<td>X. Appendix D: SOAR Ideal Materials</td>
<td>76</td>
</tr>
<tr>
<td>XI. Appendix E: Vocabulary Test</td>
<td>84</td>
</tr>
<tr>
<td>XII. Appendix F: Fact Achievement Test</td>
<td>86</td>
</tr>
<tr>
<td>XIII. Appendix G: Relationship Achievement Test</td>
<td>89</td>
</tr>
<tr>
<td>XIV. Appendix H: Concept Achievement Test</td>
<td>90</td>
</tr>
<tr>
<td>XV. Appendix I: Skill Achievement Test</td>
<td>92</td>
</tr>
<tr>
<td>XVI. Appendix J: SQ3R and SOAR Attitudes Survey</td>
<td>93</td>
</tr>
<tr>
<td>XVII. Appendix K: Control Attitudes Survey</td>
<td>94</td>
</tr>
</tbody>
</table>
LIST OF TABLES AND FIGURES

<table>
<thead>
<tr>
<th>Table/Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1</td>
<td>Theoretical Comparison of SQ3R and SOAR</td>
<td>7</td>
</tr>
<tr>
<td>Figure 1</td>
<td>List of Studies Included in SQ3R Literature Review</td>
<td>14</td>
</tr>
<tr>
<td>Table 2</td>
<td>Achievement Means and Standard Deviations for Methods, Materials, and Groups</td>
<td>39</td>
</tr>
<tr>
<td>Table 3</td>
<td>Attitude Means and Standard Deviations for Methods, Materials, and Groups</td>
<td>42</td>
</tr>
</tbody>
</table>
Chapter One: Introduction

Imagine for a moment that you are a freshman in college attending your first class of the first semester. That is a scary moment for even the most prepared, organized, and dedicated student. But what if you are not prepared, organized, and dedicated? What if you do not possess the study skills necessary to succeed? Unfortunately, this is the position that many college students, even upperclassmen, find themselves in every year (Boylan, 2002). Compounding the problem is that many college instructors expect their students to possess the study strategies necessary to succeed in their courses (Conley, 2007). This combination of factors sometimes results in students who do poorly and in instructors who do not understand why their students do poorly (Kiewra, 2005).

The main reason that students are underperforming is that they are ill equipped to deal with the rigors of college. In fact, 84% of college faculty believe that high-school graduates are either unprepared or are only somewhat well prepared to pursue a college degree (Sanoff, 2006). This unpreparedness is not surprising because instructors do not teach students how to learn. Less than 10% of instructional time is spent teaching students study strategies (Durkin, 1979; Zimmerman, Bonner, & Kovach, 1996). Instructors spend the majority of class time covering content and neglect to equip students with the strategies necessary to learn that content. Instructors probably do not teach students how to learn because they believe strategies develop naturally and therefore do not need to be taught or simply because instructors as students were never taught strategies themselves (Kiewra, 2005). Whatever the reason, the end product is students who have not learned effective study strategies and are ill equipped for college.
Students Use Ineffective Strategies

Students are generally not taught study strategies and are left to their own devices. Unfortunately, students’ preferred strategies are ineffective. Specifically, students study strategies are deficient in four key areas: a) incomplete study materials, b) disorganized study materials, c) piecemeal learning, and d) redundant strategies.

Students create incomplete study materials. For many students, academic struggles begin with note taking. Students record incomplete notes for later review, recording only about one third of lecture (Kiewra, 1985) or text (Kiewra, DuBois, Christensen, Kim, & Lindberg, 1989) details. This is not surprising because note taking, like other strategies, is not explicitly taught to students (Baker & Brown, 1984).

Complete note taking is important because note taking is positively correlated with achievement (Baker & Lombardi, 1985; Kiewra, 1984). Moreover, note taking serves both a process and product function (Kiewra, 1985). The process of taking notes is effective because it aids attention and facilitates encoding (Mayer, 2008). The product function of note taking is important because it provides an external memory aid for review (Kiewra, 1985; Titsworth, 2001). Incomplete notes are especially problematic because students have only a 5% chance of recalling non-noted information on a test (Howe, 1970).

Students create disorganized study materials. One mistake that students often make is studying directly from the notes they took during lecture or text learning. Most times these notes lack organization (Kiewra, 2005) or are organized linearly in an outline (Gubbels, 1999). The problem with disorganized or linearly organized notes is that they
make it difficult to compare across topics, such as various types of clouds in a weather unit. Graphic organizers, such as hierarchies and matrices that display information spatially, can solve the linear organization problem. Relative to outlines, graphic organizers lend to spotting (Robinson & Kiewra, 1995) and learning (Kiewra, 1994; 2005) relationships more effectively.

**Students engage in piecemeal learning.** The third problem is that students study using a piecemeal approach. They try to learn one piece of information at a time, rather than associating the pieces (Jairam & Kiewra, 2009). Working memory is capable of handling only 3-4 items at a time (Dempere-Marco, Melcher, Deco, & Kiebel, 2012); therefore, learning many separate pieces individually rather than chunking them can lead to cognitive overload (Schuler, Scheiter, & van Gunechten, 2011).

This one piece at a time “flashcard approach” to learning leads students to miss the big picture – the relationships among ideas (Kiewra, 2005). For example, a piecemeal approach might result in memorizing the height of stratus clouds, but it does not reveal how this elevation relates to weather patterns or how the height of stratus clouds relates to the heights of nimbus and cumulus clouds.

**Students rely on redundant strategies.** The problems associated with incomplete and disorganized study materials and with piecemeal learning are compounded when students use repetitive strategies to study those materials (Rachal, Daigle, & Rachal, 2007). Repetitive strategies such as rereading, reciting, rewriting, and recopying are ineffective (Anderson, 1995; Kiewra & DuBois, 1997). Even though redundant study strategies are ineffective (Callender & McDaniel, 2009), 50% of U.S.
college students typically employ them anyway (*Chronicle of Higher Education*, 2002). In fact, half of students typically study by passively reciting notes verbatim (Gubbels, 1999). Many students mistakenly believe that this increased exposure to information, in the form of reciting or rewriting notes word for word, improves performance (Jairam & Kiewra, 2009).

**Strategy Solutions**

We know that students use several ineffective strategies and that instructors are unwilling or unable to help them. The remedy, then, is not one single strategy, but a system of strategies that can address students’ widespread study problems.

This thesis looks at two complete study methods: one that is popular and long-standing and one that is contemporary and promising. These study methods are SQ3R and SOAR, respectively.

**SQ3R: Survey, Question, Read, Recite, Review.** SQ3R is a study system that has been around for decades and is comprised of five steps: Survey, Question, Read, Recite, and Review (Robinson, 1941). Robinson and other SQ3R users (Feldt & Hensley, 2009; Jacobowitz, 1988; Manzo & Manzo, 1995; Stiles, 1963) provide the following description of its steps. In the first step, Survey, students skim text headings to get an overview of the text. Students then write a statement predicting what the text is about. In the next step, Question, students turn each of the text headings into a written question. The third step of the system is Read. Here, students read the text and write an answer to each of their questions. In the fourth step, Recite, students recite the answers to each question from memory. Students do this first in their minds and then write their answers
(this is the second time that students write the answers to the questions they created). In the fifth and final step, Review, students reflect upon their questions and answers. Students consider if their questions are adequate and appropriate and if their answers are correct.

Although SQ3R is popular and has withstood the test of time, its effectiveness is questioned (Gersten, Fuchs, Williams, & Baker, 2001). It is difficult to learn and to use (Bailey, 1988; Flippo & Caverly, 2000), and arguments for its effectiveness lie mainly in opinion rather than empirical research (Johns & McNamara, 1980).

**SOAR: Select, Organize, Associate, Regulate.** The SOAR system is composed of four steps: Select, Organize, Associate, and Regulate (Kiewra, 2005; 2009). In the first step, Select, students write a set of notes that provides them with complete study materials. In the second step, Organize, students convert notes into charts or other graphic organizers that make relationships within the information apparent (Kiewra, 2005; 2009). There are four main types of organizers described by Kiewra (2005): hierarchies, sequences, matrices, and illustrations. Hierarchies reveal superordinate/subordinate relationships, sequences reveal stepwise relationships, matrices reveal comparative relationships, and illustrations reveal positional relationships. In the third step, Associate, students relate noted ideas to each other (associations within the material) and to previous knowledge (associations outside the material). In the last SOAR step, Regulate, students construct and answer potential test questions that cover various learning outcomes such as facts, relationships, concepts, and skills. It is important that students be prepared to
answer all types of questions because a test might assess any or all such learning outcomes (Kiewra, 2005).

**Theoretical Comparison of SQ3R and SOAR**

The theoretical basis for comparing SQ3R and SOAR is the information processing model, which has four basic processes (Newell & Simon, 1972): attention, encoding, storage, and retrieval. Attention involves allocating cognitive resources to the task at hand (Bruning, Schraw, Norby, & Ronning, 2004). Attention (focusing on something in particular, like a text) requires conscious effort and is limited because attention can be allocated to just one thing at a time (Anderson, 2005). When a stimulus is given attention, it might next move to long-term memory if encoding occurs (Moreno, 2010). Encoding is facilitated by incorporating and connecting new information to information already stored in long-term memory. Once encoded, information is stored in long-term memory. Information stored in visual or graphical form is stored more economically (Mayer, 2008) and is better retrieved than information stored verbally (Mayer, 1994). Last, retrieval is the process of recalling and retrieving information from long-term memory so that it can be used, such as when recalling a person’s name or the answer to a test question (Moreno, 2010).

Next I compare SQ3R and SOAR relative to each process. Table 1 is a graphic organizer summarizing the theoretical comparison of SQ3R and SOAR.
Table 1. Theoretical Comparison of SQ3R and SOAR

<table>
<thead>
<tr>
<th></th>
<th>SQ3R</th>
<th>SOAR</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Attention</strong></td>
<td>Attention is guided by Survey and Question steps, but is limited to text heading information</td>
<td>Select step aids attention through complete note taking</td>
</tr>
<tr>
<td><strong>Encoding</strong></td>
<td>No step to facilitate encoding: advocates piecemeal learning</td>
<td>Association step aids encoding by relating ideas to each other</td>
</tr>
<tr>
<td><strong>Storage</strong></td>
<td>No step for reorganization of linear notes</td>
<td>Organization step results in graphic organizers that produce economical storage</td>
</tr>
<tr>
<td><strong>Retrieval</strong></td>
<td>Self-questioning aids retrieval, but ignores different learning outcomes. Relies upon redundant strategies</td>
<td>Regulation step results in generating and answering questions for all possible learning outcomes (fact, relationship, concept, skill)</td>
</tr>
</tbody>
</table>

**Attention.** As shown in the Attention row in Table 1, both study methods aid the attention process. In SQ3R, attention is guided by the Survey and Question steps. Students attend to headings found in text and create questions based on those headings. However, students might miss important details in the text that are not directly associated with a heading. Therefore students using SQ3R might pay attention to just a subset of material (Anderson & Armbruster, 1982). The Select step of SOAR advocates that students record notes. The process of note taking, as discussed earlier, aids attention (Mayer, 2008). Moreover, the Select phase of SOAR advocates that students attend to all text ideas (Kiewra, 2005). In this regard, SOAR is superior to SQ3R because the former ensures that attention is focused on all text ideas, not just those associated with headings.
**Encoding.** As shown in the Encoding row in Table 1, only SOAR addresses the process of encoding by creating associations within the new material being learned and between the new material and prior knowledge (Mayer, 2008). Associating information alleviates the strain on working memory, diminishes cognitive load, and leads to better learning than piecemeal techniques (Anderson, 1995). Students using SQ3R do not attempt to create associations or spot relationships within the material learned or between the material and previous knowledge – they are simply reciting answers to each question. Therefore, each answer or piece of information must be encoded individually. Piecemeal approaches like this put strain on working memory (King, 1992).

**Storage.** As shown in the Storage row in Table 1, SOAR has the theoretical edge over SQ3R with regard to storage. SOAR advocates studying graphic organizers that show the relationships (i.e., hierarchical, stepwise, comparative, and positional) among ideas and make storage more economical (Mayer, 2008). SQ3R offers no means for reorganizing text that commonly follows a linear organization (Gubbels, 1999).

**Retrieval.** As shown in the Retrieval row in Table 1, both systems involve self-questioning to aid information retrieval. However, SOAR has a theoretical advantage over SQ3R because it includes testing for different learning outcomes (fact, relationship, concept, skill). Creating and answering different types of questions eliminates redundant strategies and might also improve recall of information on an achievement test (Frase & Schwartz, 1975). SQ3R contains a self-questioning component that is reliant upon redundant strategies.
Empirical Comparison of SQ3R and SOAR

SOAR has distinct theoretical advantages over SQ3R, but, to date, only one study has compared the two systems head to head. That study (Jairam, Kiewra, Rogers, Patterson-Hazley, & Marxhausen, 2012) compared the achievement outcomes of the SOAR and SQ3R systems. Participants were trained in their respective study system and then studied a text about schedules of reinforcement. In addition, students were given supplemental study materials (either SOAR or SQ3R) that matched their training condition. Following the study period, participants took a 30-item achievement test measuring fact, relationship, and concept learning. Results showed SOAR was superior to SQ3R. Students who used SOAR learned 14% more facts, 20% more relationships, and 13% more concepts than students who used SQ3R (Jairam, Kiewra, Rogers, Patterson-Hazley, & Marxhausen, 2012).

The study by Jairam et al. (2012) marked a first step in researching how SOAR and SQ3R compare. However, the study had several limitations. First, it did not examine the independent effects of training because participants trained in a study method were also given ideal materials (consistent with their training method) to study. Second, the study did not include a control group to see how SOAR and SQ3R compare with students’ preferred study methods. Third, the training for the SOAR group did not include training in concept learning, which is a vital component of the Regulate step of the SOAR method. Fourth, the study did not measure skill learning because the text material did not lend itself to skill learning. It is necessary to understand the impact of study methods on skill learning because skills represent an important learning outcome in
many content areas (e.g., science, mathematics). Fifth, the study did not assess student attitudes about the different study systems. Understanding student attitudes is important because positive attitudes might be associated with greater use (Frenzel, Goetz, Ludtke, Pekrun, & Sutton, 2009).

The purpose of the present study, then, was to compare SOAR to SQ3R, while remedying the limitations of the previous study. First, the present study examined the effects of training alone. Some trained students were not given a supplement and created their own SOAR or SQ3R materials. Second, the present study included a control group wherein students used their preferred study methods. Adding a control group allowed SOAR and SQ3R methods to be compared with students’ preferred methods. Third, training in the present study was modified to include concept training for the SOAR group. Fourth, the content for the present study was educational measurement. This new content lent itself to testing skill learning, along with fact, relationship, and concept learning. Fifth, participants filled out an attitudes survey at the conclusion of the study to provide insight into student experiences with each study method (SOAR, SQ3R, or preferred). Keeping in mind the limitations of the previous study and the modifications of the present study, the following research questions were posed:

Research Questions

1. Which study system (SQ3R, SOAR, or preferred) is best with respect to student achievement and attitudes?

2. How do SQ3R and SOAR compare, in terms of student achievement and attitudes, with training only and with both training and a provided supplement?
Hypotheses

First, it was hypothesized that SOAR studiers would outperform SQ3R studiers and preferred methods studiers across fact, relationship, and concept tests. SOAR training includes self-testing for those specific learning outcomes, whereas SQ3R and students’ preferred methods do not address specific learning outcomes. This hypothesis was also based on the theoretical advantages of SOAR over SQ3R and results from previous research showing the benefits of SOAR over students’ preferred methods (Jairam & Kiewra, 2009; 2010) and SQ3R (Jairam, Kiewra, Rogers, Patterson-Hazley, & Marxhausen, 2012). Moreover, because SQ3R can be difficult for students to learn and use (Caverly, 1985), it was hypothesized that students using SOAR would have a more positive experience than students using SQ3R.

Second, it was hypothesized that having both training and a supplement would result in higher achievement for both SQ3R and SOAR studiers than training alone. Providing students with a complete supplement should allow them to use the respective study system effectively (Kiewra, 1985) even with only a small amount of training and practice using that system.
Chapter Two: Literature Review

It is well established that students are ineffective studiers (Boylan, 2002; Kiewra, 1985; 2005). It is less clear what students and instructors can do to facilitate effective strategy use and improve learning. Two study methods have been proposed that might provide students with a complete set of strategies for academic success. These methods are SQ3R (Survey, Question, Read, Recite, Review) (Robinson, 1941) and SOAR (Select, Organize, Associate, Regulate) (Kiewra, 2005).

These two study systems were chosen for investigation because both represent a complete study method – a set of strategies for students, rather than a single strategy. A complete set of study strategies is needed to combat the many deficiencies of students’ preferred study methods (taking incomplete notes, studying disorganized or linear notes, using piecemeal learning, and relying on redundant strategies) (Kiewra, 2005). SQ3R (Robinson, 1941) was chosen for investigation because it is traditionally accepted and is routinely taught to students (Feldt & Hensley, 2009). This system has been in use for over 70 years; however, its effectiveness is questionable at best. SOAR (Kiewra, 2005) was chosen for investigation because it is a contemporary system that has shown promise in empirical studies (Jairam & Kiewra, 2009; 2010). However, its effectiveness has not been widely investigated. This literature review investigates SQ3R and SOAR and gaps in existing research.

SQ3R

Researchers often refer to SQ3R (Survey, Question, Read, Recite, Review) using terms like “time-tested” and “long-standing” (Call, 1991; Jacobowitz, 1988). SQ3R was
first introduced in the 1940s (Robinson, 1941; 1946), and its widespread presence in textbooks (Feldt & Hensley, 2009) indicates that SQ3R is well known. However, although this method has been in use for over 70 years, the research regarding its effectiveness is unclear and contradictory, as demonstrated in this literature review.

**SQ3R literature included in this review.** For the purposes of this review, 30 articles were examined, beginning with the advent SQ3R and extending to the present. A list of the included literature is found in Figure 1. Several considerations went into choosing (or excluding) literature for this review. First, due to the enormous body of SQ3R literature spanning more than 70 years, recent and historical studies alike were selected to spot trends in SQ3R research. Second, because the present study was an experimental investigation, empirical research comparing SQ3R to other strategies was therefore relevant and included. Third, because the present study pertained to helping students learn, literature involving SQ3R instruction and implementation was relevant and included. Fourth, most variations of SQ3R were excluded. Although widespread, they often bear only a slight resemblance to SQ3R (Sakta, 1999) and their inclusion might detract from the true point of interest, which is SQ3R itself. An exception is research by Walker (1991). This study is included because the variation, SRQ2R, and SQ3R are compared head to head. In all, the SQ3R literature review resulted in the inclusion of three main categories of literature: empirical studies, applied literature, and reviews.
Figure 1. List of Studies Included in SQ3R Literature Review

**Empirical Studies:**

- *SQ3R vs. Preferred Methods/Control*
  - Donald (1967)
  - Scappaticci (1977)
  - Butler (1983)
  - McCormick & Cooper (1991)
  - Carlston (2011)

- *SQ3R vs. Single Strategies*
  - Willmore (1966)
  - Snyder (1984)

- *SQ3R vs. Other Study Systems*
  - Walker (1991)
  - Jairam, Kiewra, Rogers, Patterson-Hazley, & Marxhausen (2012)

**Applied Literature:**

- *To Aid Instructor Understanding of SQ3R*
  - Robinson (1946)
  - Stiles (1963)
  - Bailey (1988)
  - Jacobowitz (1988)
  - Manzo & Manzo (1995)

- *To Aid Instructor Implementation of SQ3R*
  - Casebeer (1968)
  - Tadlock (1978)
  - Wood (1986)
  - Call (1991)
  - Feldt, Byrne, & Bral (1996)
  - Sakta (1999)
  - Feldt & Hensley (2009)
Empirical research. The empirical research includes three types of SQ3R studies:
a) studies comparing SQ3R with preferred methods, b) studies comparing SQ3R with
single strategies, and c) studies comparing SQ3R with other study systems. A list of this
literature is found in the top portion of Figure 1.

Empirical research made up approximately one third of the SQ3R research that
was reviewed. Over half of these studies compared SQ3R to control groups (Butler,
Two studies (Snyder, 1984; Willmore, 1966) compared SQ3R to the single strategies of
underlining and outlining and two (Jariam, Kiewra, Rogers, Patterson-Hazley, &
Marxhausen, 2012; Walker, 1991) compared SQ3R to other complete study systems.

SQ3R vs. preferred methods/control. SQ3R is a text-based study system and
therefore its effectiveness depends largely on reading comprehension and ability (Bailey,
1988). McCormick and Cooper (1991) tested this assertion by investigating the
effectiveness of SQ3R for students with low reading ability; in this case, students in
grades 10-12 but who read at a 6-8 grade reading level. Reading material came from
course textbooks appropriate to the participants’ reading level. The effectiveness of SQ3R was investigated three ways: in the context of a teacher-directed lesson, with students reading and studying by themselves, and with direct teacher instruction and feedback throughout the entire lesson. Achievement was assessed through verbal recall of text information. In all cases (SQ3R with teacher-directed lesson, students using SQ3R to read and study on their own, and SQ3R with direct teacher instruction throughout), SQ3R failed to improve literal comprehension compared with students who read and studied by themselves (McCormick & Cooper, 1991). This study demonstrated that students with low reading ability may struggle to use SQ3R, even when given consistent guidance and feedback from an instructor.

Critics of SQ3R have stated that SQ3R is time consuming (Feldt, Byrne, & Bral, 1996) and difficult (Flippo & Caverly, 2000) for students to learn and use. To counter that, some experimental research studies (Carlston, 2011; Donald, 1967) attempted to gauge the effectiveness of SQ3R when implemented over a long period of time. One example is a one-year study by Donald (1967) who investigated SQ3R in a 7th grade history class compared to a control group using traditional methods (e.g., group work, oral and written reports, silent and oral reading of content and answering of questions). The SQ3R group received developmental lessons in using the SQ3R steps as part of regular reading instruction. The SQ3R group also made use of “motivational devices” (p. 34) that included posters to better understand SQ3R concepts, articles stressing the purpose and value of SQ3R, and occasional questionnaires assessing pupil attitudes toward it. No motivational tools were used for the control group. Both groups had the
same instructional time and study time each day, followed the same curriculum and used the same text. Results showed that standardized test gains over the course of the year were the same for both groups. Performance on teacher-constructed tests (no additional details given) for geography and history, however, showed that the SQ3R group learned more facts than the control group. Uncontrolled variables, however, might have mitigated findings. The SQ3R and control groups had different instructors, so it is possible that the SQ3R group simply had a more effective instructor. Also, no motivational devices were employed for the control group. These types of motivational tools are not technically part of the SQ3R method; therefore, any effect they might have had cannot be attributed to the SQ3R system itself. Last, the lack of description of the teacher-constructed test makes it unclear just what was measured.

A more recent study also investigated the effectiveness of SQ3R versus preferred strategies over an extended period of time. Carlston (2011) implemented SQ3R in an introductory psychology course in the form of “note packets” over the course of the semester for seven semesters. At the beginning of each semester, students learned how to use SQ3R. The instructor (Carlston) taught the steps of SQ3R, modeled them, and provided feedback following student practice. Students were also informed that SQ3R increases exam performance. The author, however, notes that SQ3R research is anecdotal and lacking vigor, so it is interesting that he makes the opposite case to his students. Following training, the instructor informed students that SQ3R implementation and completion of a hand-written note packet demonstrating SQ3R implementation were optional but would earn extra credit. In order to be considered complete, note packets had
to include student-generated questions and answers based on headings, subheadings, identified keywords, and visual aids. Students were instructed to use these questions and answers to study for exams. Two exams were given in the course; therefore, two separate note packets were recorded, one for each unit. The instructor collected the note packets and evaluated them for completeness following each exam. Dependent measures were scores on the two course exams. No demographics were reported.

Upon completion of the course, Carlston (2011) split students into four groups: consistent completers (note packets for both tests); consistent non-completers (no note packet for either test); initiators (completed packet for second test but not first); and desistors (completed packet for first but not second). Results showed that performance from Test 1 to Test 2 increased for consistent completers and initiators, remained constant for consistent non-completers, and decreased for desistors.

Although results favored using SQ3R over one’s preferred strategies, this study does not make a convincing case for its use. The author claims that the students used SQ3R, but he rarely uses that term when reporting results and instead refers to student behavior as note-packet completion rather than SQ3R usage. This difference in terminology highlights a problem with the study – students are clearly using the Survey and Question portions of SQ3R when completing note packets, but nothing is said about the way students are studying their questions and answers. It might be that they are not truly or fully implementing the SQ3R system and that the positive results are simply attributable to an increase in student motivation (via extra credit) to take complete notes. Research, in fact, has shown a positive correlation between note taking and achievement
(Baker & Lombardi, 1985), so it is possible that it was more complete note taking and not the complete use of SQ3R that accounts for reported results.

Overall, across reviewed studies comparing SQ3R to students’ preferred methods or control groups (see Figure 1), SQ3R improved student performance only 20% of the time. However, both studies that favored SQ3R (Carlston, 2011; Donald, 1967) had methodological problems (i.e., note packets not equivalent to SQ3R and uncontrolled variables, respectively) that might mitigate results.

**SQ3R vs. single strategies.** Two experimental studies (Snyder, 1984; Willmore, 1966) that were reviewed compared SQ3R to the single strategies of underlining and outlining (see the top portion of Figure 1). In both studies, using SQ3R did not improve student performance beyond that of using a single strategy. Willmore (1966) found underlining more effective than SQ3R, whereas Snyder (1984) found both underlining and outlining comparable to SQ3R.

**SQ3R vs. other study systems.** Two experimental studies were reviewed that compared SQ3R to other study systems (as shown in the top portion of Figure 1). These systems were SRQ2R (Walker, 1991), and SOAR (Jairam, Kiewra, Rogers, Patterson-Hazley, & Marxhausen). In both studies, SQ3R was outperformed. Walker (1991) compared the effectiveness of SQ3R and a variation called SRQ2R (Survey, Read, Question, Recite, Review) on text learning for 5th grade students. The variation was based on the belief that the Question step is inappropriately placed in SQ3R (Walker, 1991). The experiment compared three groups: SRQ2R, SQ3R, and a control group. Both treatment groups were given 60 minutes of daily training in their respective system for 6
days. The type of instruction was not specified. Following training, all groups read a passage and were tested twice: one day after reading and again two weeks later. Tests consisted of short answer/essay items. Results showed that students using SRQ2R outperformed students in the SQ3R or control groups on both tests (Walker, 1991). A complete review of SOAR’s advantage over SQ3R (Jairam, Kiewra, Rogers, Patterson-Hazley, & Marxhausen, 2012) is found in the SOAR section of this literature review.

Summary of empirical SQ3R research. Many reviewers of SQ3R literature have concluded that empirical research and evidence for SQ3R is lacking (Huber, 2004; Johns & McNamara, 1980) and that SQ3R generally does not lead to higher achievement than students’ preferred methods or other strategies (Flippo & Caverly, 2000; Gersten, Fuchs, Williams, & Baker, 2001). My own analysis of empirical SQ3R research is consistent with these reviews.

Applied literature. The applied literature made up over one third of the literature reviewed and is split into two types: a) theoretical explanations of SQ3R to help instructors understand SQ3R and its steps; and b) practical descriptions and instructions aimed at helping instructors implement SQ3R into the classroom. A list of reviewed literature is found in the center portion of Figure 1.

To aid instructor understanding of SQ3R. Several articles were reviewed specifically aimed toward helping instructors understand the theoretical basis of SQ3R and the steps involved (see the center portion of Figure 1). Robinson (1946) and Stiles (1963), the oldest studies included in the review, simply outlined and described the SQ3R steps to instructors. They base their endorsement of SQ3R on theory (Robinson, 1946)
and anecdotal evidence of student improvement using SQ3R (Stiles, 1963). However, empirical evidence is not provided to substantiate claims that SQ3R improves performance. Even some of the newer literature geared toward educating instructors about SQ3R (Bailey, 1988; Jacobowitz, 1988) continues to extol the virtues of SQ3R without providing evidence for its effectiveness. Manzo and Manzo (1995), on the other hand, explain the steps of SQ3R but are critical of its effectiveness and note that students using SQ3R often do not outperform students using preferred methods.

To aid instructor implementation of SQ3R. Several articles focused on helping instructors implement SQ3R in the classroom (see the center portion of Figure 1). However, no clear and empirically supported recommendations for successful implementation were offered. Often the information offered to instructors is vague (i.e., include practice and feedback) (Feldt & Hensley 2009) or just random. Tadlock (1978), for example, advocated explaining to students why SQ3R works to help motivate them to use it. Some information aimed at instructors contends that students might have difficulty using SQ3R due to inadequate instruction on how and when to use it (Feldt & Hensley, 2009; Lipson & Wixson, 2003), but remedies for these issues are not offered. Others (Casebeer, 1968; Wood, 1986) give suggestions to instructors for implementing SQ3R in different content areas (i.e., poetry and writing), but their instructions are vague and unsubstantiated.

Information from my own review of applied research and the reviews of others (Huber, 2004; Johns & McNamara, 1980) indicates that a lack of empirical research makes it difficult to supply instructors with consistent and specific empirically supported
guidelines for SQ3R implementation. A review by Caverly (1985) can only confirm that students often have difficulty learning and applying SQ3R. And, Stahl (1984) simply offers that practice is necessary, but is not able to specify how to practice or how much practice is enough.

Summary of applied literature. A trend was observed in the reviewed applied literature: older articles (Bailey, 1988; Jacobowitz, 1988; Stiles, 1963) focused more on theoretical understanding of SQ3R, whereas newer articles (Feldt, Byrne, & Bral, 1996; Feldt & Hensley, 2009; Lipson & Wixson, 2003) focused more on instructional implementation of SQ3R, even though empirical support to merit implementation is still lacking (Huber, 2004; Manzo & Manzo, 1995).

Some early researchers advocated the use of SQ3R to teachers because they believed it was based on sound theoretical principles (Stiles, 1963; Tadlock, 1978). However, SQ3R relies upon piecemeal learning and redundant strategies, which are theoretically weak strategies. Piecemeal learning puts a strain on working memory (King, 1992) and diminishes relationship learning (Kiewra, 2005); meanwhile redundant strategies are ineffective (Anderson, 1995; Callender & McDaniel, 2009) and do not improve performance (Jairam & Kiewra, 2009). In my review of applied literature, authors occasionally addressed the lack of empirical support for SQ3R (Huber, 2004; Manzo & Manzo, 1995), but never addressed the theoretical shortcomings of SQ3R. And no articles offered systematic and empirically supported approaches to effective SQ3R classroom implementation.
**Reviews.** The general consensus of included reviews is that research (see the bottom portion of Figure 1) regarding SQ3R is inconclusive (Gersten, Fuchs, Williams, & Baker, 2001) and that empirical evidence for the effectiveness of SQ3R is lacking (Huber, 2004). Many of the included reviews specifically point out that SQ3R might be difficult for students to learn or apply, particularly without sufficient practice (Caverly, 1985; Flippo & Caverly, 2000; Stahl, 1984).

Some researchers believe that one strong point of SQ3R is that students can use it independently (Tadlock, 1978). Conversely, the review by Huber (2004) concluded that SQ3R is most effective when used under the following conditions: students must be informed of its purpose, SQ3R must be logically related to the text materials, students must receive feedback, and students must be given sufficient amounts of guided and independent practice using SQ3R (although what counts as sufficient is not specified). These conditions suggest that students new to SQ3R might have difficulty using it effectively on their own. Unfortunately, in most of the experimental research, it is difficult to determine how much SQ3R training and practice students were given. A lack of sufficient training and practice might contribute to decreased effectiveness of SQ3R in experimental settings.

SQ3R has retained the support of some researchers, but the reasons given for that support are largely theoretical and opinion-based rather than empirical. One example is the reasoning Jacobowitz (1988) provides for why SQ3R is the best option for students: it is supported by cognitive theory, it is mentioned in many secondary and college level study skills texts, it is a popular skill that has been in use for a long time, and it is highly
structured and easy to learn. Jacobowitz, though, offers no empirical support to back these claims.

**Gaps in research.** Even with the obvious lack of empirical support for SQ3R, it continues as a mainstay in the canon of study strategies (Feldt & Hensley, 2009). Although SQ3R seems to provide a systematic approach to studying, the following gaps in SQ3R research are difficult to overlook.

**Overall lack of current, methodologically sound research.** The vast majority of support for SQ3R comes from its reputation and anecdotal evidence. Due to this, many have simply accepted it as effective and continue to include it in textbooks, teach it in classrooms, and extol its virtues as a study method (Feldt & Hensley, 2009). Unfortunately, the lack of empirical evidence is troubling. Even Robinson (1946), who devised the model, could not provide any empirical evidence for the effectiveness of SQ3R (Johns & McNamara, 1980).

Additionally, the majority of experimental research on SQ3R is more than 20 years old. The past couple of years have seen a small resurgence of interest in SQ3R (Carlston, 2011; Jairam, Kiewra, Rogers, Patterson-Hazley, & Marxhausen, 2012), perhaps because researchers are realizing how little we actually know about its effectiveness.

**Lack of research comparing SQ3R to complete study methods.** Although SQ3R has been compared to single strategies like underlining and outlining (Snyder, 1984; Willmore, 1966), there is a lack of research comparing it to complete study systems, even to variations that it inspired. One study has compared SQ3R to SRQ2R (Walker, 1991)
and one has compared SQ3R to SOAR (Jairam, Kiewra, Rogers, Patterson-Hazley, & Marxhausen, 2012) (this latter study is reviewed in depth in the SOAR section of the literature review), and in both cases SQ3R was outperformed. More research is needed to gauge the effectiveness of SQ3R compared to other complete study methods to determine which study method works best for students.

In general, although SQ3R is well known, its limited empirical research demonstrates that it is largely ineffective (Butler, 1983; Huber, 2004; McCormick & Cooper, 1991). Skepticism about the effectiveness of SQ3R has increased over the years. Early research (Stiles, 1963) showed more support for the system (based on theory and anecdotal evidence), whereas more recent research has been more critical of its effectiveness (Flippo & Caverly, 2000; Huber, 2004).

SOAR

SOAR (Select, Organize, Associate, Regulate) is a much more contemporary system (Kiewra, 2005) than SQ3R, and therefore does not yet have the abundant literature base of SQ3R. However, SOAR is based on sound theoretical principles, and SOAR research appears methodologically sound and shows promising results.

Types of literature. The existing SOAR research falls into two main categories: research examining the use of SOAR as an integrated study system and research that supports theoretically the component parts of SOAR.

Research on SOAR study system. Three major studies have been conducted examining the effectiveness of SOAR as a complete study system. The first (Jairam & Kiewra, 2009) compared the additive effects of SOAR steps for text learning among five
groups: Select (S) only; Select and Organize (SO); Select, Organize, and Associate (SOA); Select, Organize, Associate, and Regulate (SOAR); and a control group using preferred methods. In this study, students were not trained in how to use SOAR. Instead, they were given their respective SOAR materials to help them learn information from a text about wildcats: the S group was given a complete set of notes organized in linear form; the SO group was given a complete set of notes organized into a matrix; the SOA group studied the same matrix along with a list of 27 associations; and the SOAR group studied the matrix and associations plus regulation practice questions (with provided answers). The control group studied the same wildcat text as the other groups but was not given additional study aids. In general, those using the full SOAR system outperformed those using just some of its parts or preferred methods (Jairam & Kiewra, 2009).

The second study involving SOAR was also conducted by Jairam and Kiewra (2010) and was an extension of the first. The groups remained the same; however, the same wildcat materials were computer-based rather than text-based. Again, no specific training in the SOAR system was provided, but this time students assisted in creating SOAR materials on the computer. Students in the S group completed a blank linear framework by clicking on facts found in the text. Students in the SO group did the same but inserted facts within a matrix framework. SOA group materials were identical to SO materials with the addition of a list of provided associations that became available upon completion of the matrix. When students clicked on these associations, they were highlighted in the completed matrix. The SOAR group materials were the same as the SOA group materials with the addition of provided practice questions. Students clicked
on possible answers and were given feedback on their correctness. Students in the control group created their own set of study materials using their preferred methods. Results mirrored those for Jairam and Kiewra (2009). Achievement on fact and relationship tests generally increased with the number of SOAR strategies used (SOAR > SOA > SO > S > Control) (Jairam & Kiewra, 2010).

The third and final study compared SOAR head-to-head with SQ3R (Jairam, Kiewra, Rogers, Patterson-Hazley, & Marxhausen, 2012) and did not include a control group. Training in either SOAR or SQ3R was provided through the use of Powerpoint slideshows, but students were not asked to create their own SOAR or SQ3R study materials. Rather, experimenter-created supplements meant to reflect ideal usage of the respective system were provided. Results revealed that students who used SOAR performed significantly higher on both fact and relationship achievement tests (and marginally higher on a concept test) than students using SQ3R (Jairam, Kiewra, Rogers, Patterson-Hazley, & Marxhausen, 2012).

**Research supporting SOAR components.** Each step of SOAR was designed to address and correct a common strategy error by students, and each of the corrective strategies is empirically supported. The following section provides a brief snapshot of research regarding each component of the SOAR system.

The Select step is designed to correct incomplete note taking by encouraging students to create a complete set of study materials. Complete notes are imperative because note taking is positively correlated with achievement (Baker & Lombardi, 1985). Research shows that students who take notes from text or from lecture learn more than
students who do not take notes, even when those notes are not reviewed (Kiewra, 1987; 1988). The act of note taking aids attention (Crooks, White, & Barnard, 2007) and encoding (Kiewra, 1985). Notes are most valuable, though, because they provide an external record for review (Kiewra, 1985). Due to empirical support for note taking, including it in a complete study system makes sense.

There is also empirical support for including an organization component. The Organize step of SOAR corrects the disorganized or linearly organized materials problem by employing the use of graphic organizers. Students typically organize notes linearly into lists or outlines (Gubbels, 1999); however, research shows that studying information in matrix form rather than linear form improves achievement (Robinson & Kiewra, 1995). Linear organization especially inhibits relationship learning compared to graphic organization (Robinson & Kiewra, 1995). Organized information is more easily encoded and stored in long-term memory than disorganized information (Mayer, 2008).

The Associate step addresses piecemeal learning by promoting associative learning. Association has been known alternately as generative learning (Wittrock, 1990), elaboration (Anderson, 1995), and integration (Mayer, 1996). Essentially, association works by linking pieces of information to each other and to information students already know. Not only does association facilitate encoding (Mayer, 2008), it also improves understanding of the material and aids in the retrieval process (Mayer, 1996). Additionally, research shows that students learn more when using associative strategies than when using piecemeal strategies (King, 1992).
The Regulation step corrects student usage of redundant strategies by promoting self-testing. Self-testing improves learning, although students often use it only to gauge their learning, not to enhance it (Karpicke, Butler, & Roediger, 2009). Regardless, research shows that students who generate and answer practice test questions recall more information than students who do not include self-testing during the review process (Frase and Schwartz, 1975). Therefore, the Regulation step not only helps students to gauge their learning, but also to improve it.

**Summary.** SOAR is a complete study system based on sound theoretical principles (see thesis introduction) and empirical research. Specifically, it has shown promise when compared to its component parts and to students’ preferred methods (Jairam & Kiewra, 2009; 2010) and to the SQ3R study system (Jairam, Kiewra, Rogers, Patterson-Hazley, & Marxhausen, 2012). This latter study, though, is the only one that has compared SOAR and SQ3R and that study has limitations that leave research gaps to be filled.

**Gaps in research.** Although the three studies (Jairam & Kiewra, 2009; 2010; Jairam, Kiewra, Rogers, Patterson-Hazley, & Marxhausen, 2012) that have examined SOAR as an integrated system are a good start, there are some gaps that exist. First, none of the SOAR studies examine the effectiveness of SOAR when students create their own materials. In all cases, students are presented with experimenter-created materials or partially created materials. Second, research has demonstrated that SOAR improves fact and relationship learning (Jairam & Kiewra, 2009; 2010), but it is unclear how SOAR impacts concept and skill learning. Third, SOAR has not been tested across a broad range
of content areas, so it is difficult to determine how effective SOAR might be across different content areas, especially those requiring skill learning.
Chapter Three: Methods

Participants and Design

One hundred thirty-eight undergraduate students enrolled in educational psychology classes at a large Mid-western university volunteered to participate in this study. Approval to conduct this study was granted by the University’s Institutional Review Board prior to data collection (IRB #20120512514 EX). All students who participated received course credit. Participants ranged from freshmen to seniors, although most were sophomores (44%) and juniors (30%). Participants’ GPAs ranged from 2.5 to 4.0, although 80% had a GPA higher than 3.0. Participants were 86% female and 14% male. Participants were assigned randomly to one cell of a 2x2 factorial design or to a control group. The two factors were study method (trained in SQ3R or SOAR) and study material (received supplement (SQ3R or SOAR, respectively) or no supplement). The Control group studied using their preferred methods.

Materials

Materials included a demographic survey, as well as materials for the different phases of the experiment: training, studying, and testing.

The demographic survey (see Appendix A) was comprised of six multiple-choice items that determined: (1) gender, (2) class standing, (3) overall GPA, (4) prior experience with SOAR, (5) prior experience with SQ3R, and (6) previous knowledge of the study material (educational measurement).
**Training phase.** The training phase consisted of a Powerpoint slideshow for each method (SOAR, SQ3R, or control). Each slideshow was tailored to provide training in the respective method. The slideshows were adapted from a previous experiment conducted by Jairam, Kiewra, Rogers, Patterson-Hazley, and Marxhausen (2012). Each slideshow was timed to advance automatically, and each took 30 minutes to complete. The slideshows were consistent across groups. Each addressed introduction of the study method, an example of the method, an example with guided practice for the student, and finally an example with unguided practice. The only difference in the training slideshows was the nature of the training itself. Participants in the SOAR group were presented with a slideshow explaining and demonstrating the four steps of the SOAR study system (Select, Organize, Associate, and Regulate). Participants in the SQ3R group were presented with a slideshow explaining and demonstrating the five steps of the SQ3R system (Survey, Question, Read, Recite, and Review). Participants in the control group were asked to use their preferred study methods to complete the training. All participants were provided a notepad for note taking.

During the introduction phase of training, the steps of each system were listed and explained. Next, participants were presented with an example of the system being used for a short passage dealing with symbiosis. For the guided example about animal behavior, each participant was prompted to practice the study methods step-by-step using the provided notebook. Participants were given feedback and examples following each step. Last, during the unguided practice example, participants were asked to use the entire system, without prompts, for a passage about wildcats. Feedback for all steps was
provided after the completion of the unguided example.

**Studying phase.** The materials for the study phase included a text covering educational measurement and an additional blank notepad for the purpose of taking notes. Participants in the supplement groups were also provided with a supplement of either ideal SOAR or SQ3R materials, respectively. The educational measurement text was approximately 2500 words and presented on six standard typed and printed pages (see Appendix B). It addressed the three measures used to describe sets of scores: central tendency (mean, median, and mode), dispersion (range, variance, and standard deviation), and shape of the distribution (normal, skewed, and bimodal). For each topic, the following information was covered, but without headings: definition, use, and calculation. In addition, the text contained appropriate figures and examples.

The SQ3R and SOAR supplemental materials were created by the experimenter and were intended to represent the most ideal usage of each system, respectively. The SQ3R ideal materials (see Appendix C) consisted of a brief overview of the text (Survey); thirteen written questions created from text headings (Question); complete answers to each of those questions, as found in the text (Read); a restatement of questions and answers (Recite); and an assessment of the appropriateness of the questions and the accuracy of answers (Review).

The SOAR ideal materials (see Appendix D) contained a set of linear notes in bullet-like format (Select). To aid Organization, the following graphic organizers were provided: a hierarchical organizer showing the three basic measures used to describe sets of scores and their sub-measures, a matrix organizer comparing the three types of central
tendency (with regard to definition, calculation, affect of extreme scores, and complexity), a matrix organizer comparing the three types of dispersion (with regard to definition, calculation, complexity, and limitations), and a matrix organizer comparing the different shapes of distributions (with regard to description; illustration; placement of mean, median, and mode; and most appropriate measure of central tendency). To aid Association, a list of nine associations from within the material and seven associations outside the material were provided. And to aid Regulation, a 24-question quiz comprised of 13 fact, 5 relationship, and 6 concept questions, with answers to each provided on the following page, was provided.

**Testing phase.** Two tests were used, vocabulary and achievement, along with an attitude survey. The vocabulary test (see Appendix E) was comprised of six multiple-choice items taken from the verbal portion of a sample Scholastic Aptitude Test. The vocabulary test served as a distracter task to clear the educational measurement information from short-term memory.

The achievement tests (see Appendices F-I) included 36 multiple-choice items, each with four choices. Fifteen items measured fact learning (e.g., What is the best measure of central tendency for a negatively skewed distribution?), 5 items measured relationship learning (e.g., Which two types of central tendency are least affected by extreme scores?), 10 items measured concept learning (e.g., Mrs. Griffin found that the test scores of her students centered on the score of 87. She later found out that one of the test scores had to be changed from 68 to 0 because the student had cheated. What measure of central tendency would be most affected?), and 6 items measured skill
learning (e.g., What is the median for the following set of scores: 2, 2, 4, 5, 7, 7, 8, 8, 9). Fact, relationship, and concept test items were similar in form but differed in content from the regulation items found in the SOAR ideal materials.

The attitude survey for the SQ3R and SOAR groups (see Appendix J) contained 6 Likert-type items with four choices (1= Strongly Disagree, 2= Disagree, 3= Agree, 4= Strongly Agree) that assessed participants’ attitudes about the study system used and their research experience. The survey asked if: a) the training prepared them to study; b) the system was easy to learn; c) the system was easy to use; d) the system was effective; e) the system was enjoyable; and f) the system is one they will reuse. The survey also included an open-ended question for students to share any additional information about their study methods. Because the control group was not trained in a new system but rather trained in their preferred methods, they answered only the first survey question and the open-ended question (see Appendix K).

Procedure

All participants met in a campus computer lab and were seated randomly at computers. Each computer had previously been loaded with the training materials for the SOAR, SQ3R, or control groups. The respective Powerpoint presentations were already open on computer screens and prepared to start with the click of a mouse. All other materials were presented as hard copies and were placed in order of use on the desk of each participant. All participants were first asked to read and sign the informed consent form. Next, participants were read a set of general instructions about the research procedure (e.g., turn off cell phones, remain seated and silent, pay attention only to your
own work, use all the time allotted for tasks), completed the demographic survey, and
then progressed through the training, studying, and testing phases of the experiment.
During the 30-minute training phase, participants viewed their respective slideshow
(SOAR, SQ3R, or control) on the computer and completed the examples on the provided
paper notepad. Next, during the 45-minute studying phase, participants in the SOAR/no
supplement, SQ3R/no supplement, and control groups were given the text on educational
measurement and an additional notepad on which to create materials or take notes.
Participants in the SOAR/supplement and SQ3R/supplement groups were also given the
respective ideal materials to study along with the text. Participants were told they were
preparing for a test over the text materials and were allowed to take notes on any of the
provided materials. Last, during the testing phase, participants took the vocabulary test
and were then given approximately 20 minutes to complete the achievement test over the
educational measurement material. Finally, all participants were asked to complete the
survey about their experience, were debriefed, and dismissed. The entire experimental
procedure took two hours to complete.

Scoring

Scoring of questions 1-6 of the SQ3R and SOAR attitude surveys and question 1
of the Control attitude survey was objective and was performed by the experimenter.
Scoring of the multiple-choice achievement tests was also objective. Students recorded
their answers on Scantron forms, and these forms were computer scored. Scoring of the
open-ended survey questions and study notepads was subjective and performed by the
experimenter who looked primarily for evidence of strategy use.
Chapter Four: Results

Initial inspection of the data showed that participants had little to no experience in SOAR or SQ3R methods and had little knowledge about the educational measurement topic.

For each achievement (fact, relationship, concept, skill) and attitude (prepared to study, easy to learn, easy to use, effective, enjoyable, reusable) dependent measure, two analyses were conducted. The first was a 2x2 ANOVA that examined main effects of method (SOAR, SQ3R) and material (supplement, no supplement), as well as the method x material interaction. The second analysis was a one-way ANOVA that compared each of the four treatment groups (SOAR/supplement, SOAR/no supplement, SQ3R/supplement, SQ3R/no supplement) and the control group.

Achievement

Fact. Results from the 2 x 2 ANOVA indicated that the main effect of study method, \(F(1,88) = .04, p = .84, MSE = 6.31\), the main effect of material, \(F(1,88) = 3.81, p = .054, MSE = 6.31\), and the method x material interaction \(F(1,88) = .08, p = .77, MSE = 6.31\) had no effect on fact performance. The means for these analyses are found in the top portion of the Fact column in Table 2. It should be noted, however, that the main effect of material just missed achieving statistical significance and is worthy of attention. As can be seen in Table 2, those studying supplements learned somewhat more facts than those without supplements. This finding suggests that the combination of strategy training and supplements aids fact learning to some degree.
The result from the one-way ANOVA comparing the five groups was also not significant, $F(4, 133) = 1.22, p = .31, MSE = 5.32$. Thus, none of the treatment groups outperformed the control group. The means for this analysis are found in the bottom portion of the Fact column in Table 2.

**Relationship.** Results from the 2x2 ANOVA indicated that the main effect of study method ($F(1, 88) = .34, p = .56, MSE = 1.57$), the main effect of material ($F(1, 88) = .01, p = .93, MSE = 1.57$), and the method x material interaction ($F(1, 88) = .17, p = .68, MSE = 1.57$) had no effect on relationship performance. The means for these analyses are found in the top portion of the Relationship column in Table 2.

The result from the one-way ANOVA comparing the five groups was also not significant, $F(4, 133) = .15, p = .96, MSE = 1.40$. The means for this analysis are found in the bottom portion of the Relationship column in Table 2.

**Concept.** Results from the 2x2 ANOVA indicated that the main effect of study method ($F(1, 88) = .003, p = .95, MSE = 3.26$), the main effect of material ($F(1, 88) = .27, p = .61, MSE = 3.26$), and the method x material interaction ($F(1, 88) = .75, p = .39, MSE = 3.26$) had no effect on concept performance. The means for these analyses are found in the top portion of the Concept column in Table 2.

The result from the one-way ANOVA comparing the five groups was also not significant, $F(4, 133) = 1.18, p = .32, MSE = 2.86$. The means for this analysis are found in the bottom portion of the Concept column in Table 2.
Table 2. Achievement Means and Standard Deviations for Methods, Materials, and Groups

<table>
<thead>
<tr>
<th></th>
<th>Achievement Tests</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fact (n = 15)</td>
<td>Relationship (n = 5)</td>
<td>Concept (n = 10)</td>
<td>Skill (n = 6)</td>
</tr>
<tr>
<td><strong>Method</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SOAR (n = 46)</td>
<td>11.54 (2.80)</td>
<td>4.13 (1.19)</td>
<td>7.85 (1.76)</td>
<td>4.52 (1.23)</td>
</tr>
<tr>
<td>SQ3R (n = 46)</td>
<td>11.65 (2.24)</td>
<td>3.98 (1.29)</td>
<td>7.83 (1.83)</td>
<td>4.41 (1.13)</td>
</tr>
<tr>
<td><strong>Material</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supplement (n = 46)</td>
<td>12.11 (2.31)</td>
<td>4.07 (1.27)</td>
<td>7.93 (1.69)</td>
<td>4.65 (1.02)</td>
</tr>
<tr>
<td>No Supplement (n = 46)</td>
<td>11.09 (2.65)</td>
<td>4.04 (1.21)</td>
<td>7.74 (1.89)</td>
<td>4.28 (1.29)</td>
</tr>
<tr>
<td><strong>Group Comparisons</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SQ3R/No supplement (n = 23)</td>
<td>11.22 (2.49)</td>
<td>3.91 (1.13)</td>
<td>7.57 (2.04)</td>
<td>4.26 (1.32)</td>
</tr>
<tr>
<td>SQ3R/Supplement (n = 23)</td>
<td>12.09 (1.93)</td>
<td>4.04 (1.46)</td>
<td>8.09 (1.59)</td>
<td>4.57 (.90)</td>
</tr>
<tr>
<td>SOAR/No supplement (n = 23)</td>
<td>10.96 (2.85)</td>
<td>4.17 (1.30)</td>
<td>7.91 (1.76)</td>
<td>4.30 (1.30)</td>
</tr>
<tr>
<td>SOAR/Supplement (n = 23)</td>
<td>12.13 (2.69)</td>
<td>4.09 (1.08)</td>
<td>7.78 (1.81)</td>
<td>4.74 (1.14)</td>
</tr>
<tr>
<td>Control (n = 46)</td>
<td>11.78 (1.84)</td>
<td>4.09 (1.03)</td>
<td>8.41 (1.44)</td>
<td>4.76 (.99)</td>
</tr>
</tbody>
</table>
Skill. Results from the 2x2 ANOVA revealed that the main effect of study method \((F(1,88) = .20, p = .66, MSE = 1.38)\), the main effect of material \((F(1,88) = 2.28, p = .14, MSE = 1.38)\), and the method x material interaction \((F(1,88) = .07, p = .79, MSE = 1.38)\) had no effect on skill performance. The means for these analyses can be found in the top portion of the Skill column in Table 2.

The result from the one-way ANOVA comparing the five groups was also not significant, \(F(4,133) = 1.23, p = .30, MSE = 1.25\). The means for this analysis can be found in the bottom portion of the Skill column in Table 2.

Summary. Across all four achievement tests, SOAR and SQ3R trained groups did not differ from one another whether they were given addition study supplements or not. Moreover, none of the treatment groups differed from the preferred strategy control group. In one case, results did approach statistical significance: the supplement groups learned somewhat more facts than the non-supplement groups.

Attitudes

The attitude survey was completed by all participants. However, because the control group was not trained in a new system (SOAR or SQ3R), survey items 2-6 were not relevant to them and were therefore not administered. As such, control group participants answered only Question 1 (prepared to study) on the survey.

Question 1: Prepared to study. Results from the 2x2 ANOVA indicated that study method had a significant main effect on student attitudes for preparedness, \(F(1,88) = 5.18, p = .03, MSE = .25\). Students trained in SQ3R felt more prepared to study than students using SOAR. Neither the main effect of material \((F(1,88) = .39, p = .54, MSE = \)
nor the material x method interaction ($F(1,88) = .39, p = .54, MSE = .25$) were significant. The means for these analyses can be found in the top portion of the Prepared to Study column of Table 3.

The result from the one-way ANOVA comparing the five groups was also significant, $F(4,133) = 4.16, p = .003, MSE = .36$. Students in the SQ3R/supplement and SQ3R/no supplement groups felt more prepared to study than those in the control group. The means for this analysis can be found in the bottom portion of the Prepared to Study column of Table 3.

**Question 2: Easy to learn.** Results from the 2x2 ANOVA revealed that the main effect of study method ($F(1,88) = 2.16, p = .15, MSE = .32$), the main effect of material ($F(1,88) = .14, p = .71, MSE = .32$), and the method x material interaction ($F(1,88) = 1.22, p = .27, MSE = .32$) had no effect on how easy to learn the study systems were perceived to be. The means for these analyses can be found in the top portion of the Easy to Learn column of Table 3.

The result of the one-way ANOVA comparing the four groups was also not significant, $F(3,88) = 1.17, p = .33, MSE = .32$. The means for this analysis can be found in the bottom portion of the Easy to Learn column of Table 3.
Table 3. Attitude Means and Standard Deviations for Methods, Materials, and Groups

<table>
<thead>
<tr>
<th>Attitudes Survey</th>
<th>Prepared To Study</th>
<th>Easy to Learn</th>
<th>Easy to Use</th>
<th>Effective</th>
<th>Enjoyable</th>
<th>Reusable</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Method</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SOAR (n = 46)</td>
<td>3.07 (.49)</td>
<td>3.22 (.51)</td>
<td>3.17 (.49)</td>
<td>3.02 (.61)</td>
<td>2.59 (.58)</td>
<td>2.72 (.62)</td>
</tr>
<tr>
<td>SQ3R (n = 46)</td>
<td>3.30 (.51)</td>
<td>3.39 (.61)</td>
<td>3.26 (.61)</td>
<td>3.07 (.49)</td>
<td>2.50 (.62)</td>
<td>2.63 (.61)</td>
</tr>
<tr>
<td><strong>Material</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supplement (n = 46)</td>
<td>3.22 (.51)</td>
<td>3.33 (.47)</td>
<td>3.33 (.52)</td>
<td>3.13 (.50)</td>
<td>2.57 (.62)</td>
<td>2.72 (.62)</td>
</tr>
<tr>
<td>No Supplement (n = 46)</td>
<td>3.15 (.52)</td>
<td>3.28 (.66)</td>
<td>3.11 (.57)</td>
<td>2.96 (.60)</td>
<td>2.52 (.59)</td>
<td>2.63 (.61)</td>
</tr>
<tr>
<td><strong>Group Comparisons</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SQ3R/No supplement (n = 23)</td>
<td>3.30 (.56)</td>
<td>3.30 (.70)</td>
<td>3.13 (.63)</td>
<td>3.00 (.52)</td>
<td>2.52 (.59)</td>
<td>2.65 (.65)</td>
</tr>
<tr>
<td>SQ3R/Supplement (n = 23)</td>
<td>3.30 (.47)</td>
<td>3.48 (.51)</td>
<td>3.39 (.58)</td>
<td>3.13 (.46)</td>
<td>2.48 (.67)</td>
<td>2.61 (.58)</td>
</tr>
<tr>
<td>SOAR/No supplement (n = 23)</td>
<td>3.00 (.43)</td>
<td>3.26 (.62)</td>
<td>3.09 (.52)</td>
<td>2.91 (.67)</td>
<td>2.52 (.59)</td>
<td>2.61 (.58)</td>
</tr>
<tr>
<td>SOAR/Supplement (n = 23)</td>
<td>3.13 (.55)</td>
<td>3.17 (.39)</td>
<td>3.26 (.45)</td>
<td>3.13 (.55)</td>
<td>2.65 (.57)</td>
<td>2.83 (.65)</td>
</tr>
<tr>
<td>Control (n = 46)</td>
<td>2.80 (.75)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Attitudes scored on a 4-point scale. 1 = Strongly Disagree, 2 = Disagree, 3 = Agree, 4 = Strongly Agree
**Question 3: Easy to use.** Results from the 2x2 ANOVA indicated that the main effect of study method \( (F(1,88) = .58, p = .45, MSE = .30) \), the main effect of material \( (F(1,88) = .363, p = .06, MSE = .30) \), and the method x material interaction \( (F(1,88) = .15, p = .70, MSE = .30) \) had no effect on how easy to use the study systems were perceived to be. The means for these analyses can be found in the top portion of the Easy to Use column of Table 3.

The result of the one-way ANOVA comparing the four groups was non-significant, \( F(3,88) = 1.45, p = .23, MSE = .30 \). The means for this analysis can be found in the bottom portion of the Easy to Use column in Table 3.

**Question 4: Effective.** The results from the 2x2 ANOVA revealed that the main effect of study method \( (F(1,88) = .14, p = .71, MSE = .31) \), the main effect of material \( (F(1,88) = 2.26, p = .14, MSE = .31) \), and the method x material interaction \( (F(1,88) = .14, p = .71, MSE = .31) \) had no effect on how effective the study systems were perceived to be. The means for these analyses are found in the top portion of the Effective column of Table 3.

The result of the one-way ANOVA comparing the four groups was not significant, \( F(3,88) = .85, p = .47, MSE = .31 \). The means for this analysis can be found in the bottom portion of the Effective column in Table 3.

**Question 5: Enjoyable.** Results from the 2x2 ANOVA indicated that the main effect of method \( (F(1,88) = .47, p = .49, MSE = .37) \), the main effect of material \( (F(1,88) = .12, p = .73, MSE = .37) \), and the material x method interaction \( (F(1,88) = .47, p = .49, MSE = .37) \) had no effect on how enjoyable the study systems were perceived to be. The
means for these analyses can be found in the top portion of the Enjoyable column in Table 3.

The result of the one-way ANOVA comparing the four groups was not significant, $F(3,88) = .35, p = .79, MSE = .37$. The means for this analysis can be found in the bottom portion of the Enjoyable column in Table 3.

**Question 6: Reusable.** Results from the 2x2 ANOVA revealed that the main effect of method ($F(1,88) = .46, p = .50, MSE = .38$), the main effect of material ($F(1,88) = .46, p = .50, MSE = .38$), and the method x material interaction ($F(1,88) = 1.03, p = .31, MSE = .38$) had no effect on how likely participants were to reuse their study system. The means for these analyses can be found in the top portion of the Reusable column in Table 3.

The result of the one-way ANOVA comparing the four groups was not significant, $F(3,88) = .65, p = .59, MSE = .38$. The means for this analysis can be found in the bottom portion of the Reusable column in Table 3.

**Summary.** Across all 6 items, student attitudes were relatively the same regardless of group affiliation. The lone exception pertained to how well training prepared students to use the study methods. In that regard, those trained in SQ3R felt more prepared than those trained in SOAR and preferred methods.

**Open-Ended Survey Question Observations**

Responses to the open-ended survey question were also examined to determine if the present study had the intended experimental effects. That is, did students employ the study system they were trained in and use the provided supplements as planned? A
couple interesting observations emerged. First, among students in the SQ3R/no supplement group, approximately 90% reported that they used SQ3R to study the material, whereas the other 10% reported reverting back to preferred study methods. Among students in the SOAR/no supplement group, 75% reported using the SOAR system, whereas 25% reported reverting back to preferred methods. A large number of students in both the SQ3R/no supplement and SOAR/no supplement groups reported that they felt like they did not have sufficient time to create study materials and review them, whereas none of the students in the supplement groups reported feeling short of time. Also, several students in all treatment groups reported that it was difficult for them to break away from using their preferred methods, even if they reported that the new system (SQ3R or SOAR) was easy to use and easy to learn.

**Study Notepad Observations**

The study notepads were also examined to determine if the present study had the intended experimental effects. Observations of study notepads somewhat confirmed what students reported in the open-ended survey question regarding system usage. That is, students strived to use their trained methods but insufficient time might have prevented them from using those methods fully. This was especially true for the no-supplement groups that needed to create their study materials. Those in the SQ3R/no supplement group generally used the Survey, Question, and Read steps. However, the amount of detail in student answers varied greatly, from a single word or phrase to a short paragraph, for each question. However, it was difficult to assess from the notepads if students followed the Recite and Review steps because these might not involve a written
response. Those in the SOAR/no supplement group used the Select and Organize steps successfully. The final two steps, Associate and Regulate, were used less often or were carried out only partially. Many SOAR studiers created fewer than five associations and five regulation questions, and a few students created none of either. Notepads of students in the control group typically included linear notes, although some students also practiced educational measurement calculations.

Students in both supplement groups oddly spent time creating their own notes even though they had been given a set of complete study materials. The length of notes among those receiving supplements varied greatly. Most were one page or less; some were several pages. Upon further inspection, notepads from SQ3R/supplement users typically showed use of Survey, Question, and Read steps, although some notepads revealed students falling back to preferred strategies and simply recording linear notes. Notepads from SOAR/supplement users typically showed the Select step and attempts to answer practice questions provided in the Regulation portion of the supplement.
Chapter Five: Discussion

The Discussion includes three parts: summary of findings; study limitations and future research directions; and instructional implications and conclusions.

Summary of Findings

The purpose of this study was to determine what study system (SQ3R, SOAR, or preferred methods) is best with respect to achievement (fact, relationship, concept, and skill) and student attitudes (prepared to study, easy to use, easy to learn, effective, enjoyable, and reusable), and how SQ3R and SOAR compare with training only and with training plus a supplement.

Results showed no significant differences between study systems (SQ3R, SOAR, or preferred methods) with respect to achievement across fact, relationship, concept, or skill learning outcomes. Therefore, results did not support the main hypothesis that SOAR students would outperform students using SQ3R and students using preferred methods across fact, relationship, and concept tests. This finding does not fit with previous research showing that students using SOAR outperformed students using SQ3R for fact and relationship learning (Jairam, Kiewra, Rogers, Patterson-Hazley, & Marxhausen, 2012), or with previous research demonstrating that using SQ3R (Carlston, 2011) or SOAR (Jairam & Kiewra, 2009; 2010) leads to higher achievement than students’ preferred methods.

However, present findings did reveal that students trained in SQ3R felt more prepared to study than students trained in SOAR or preferred methods. This unexpected finding runs counter to the main hypothesis that students would have more positive
attitudes about SOAR than about SQ3R or preferred methods. This single attitude finding should perhaps be taken lightly because the groups did not differ on other attitude statements. Alternatively, the SQ3R system is simpler than SOAR and perhaps that simplicity helped students feel more prepared.

The results of this study also showed no systematic benefit of supplements. Those who received training plus SQ3R or SOAR supplements did not achieve more than those who received training and no supplements. Although the results of this study did not uniformly support the hypothesis that providing students with a supplement boosts achievement, there was one exception. The lone exception occurred for fact learning, where the analysis approached statistical significance and showed an advantage for supplements. This single finding suggests that students used supplements in a limited way—mainly for fact learning even though the SOAR supplement, in particular, was designed to aid all learning outcomes. Perhaps more importantly, the present study failed to show a training-only (without supplement) benefit. The present study is the first to examine the training benefits of SOAR alone and to compare SOAR and SQ3R training without supplements. For now, it appears that SOAR is no more effective than SQ3R when supplements are not provided as they were in previous studies by Jairam and Kiewra (2009; 2010).

It is important to try to understand why present results differed from those of Jairam et al. (2012) where SOAR was found superior to SQ3R. One factor might be the different materials across studies. Perhaps the present materials were too easy or too difficult to uncover differences. That, however, does not appear to be the case. Overall,
achievement was neither too low nor too high to indicate floor or ceiling effects. Means were between 76% and 81% for the four tests.

Some insight into comparable group achievement did come from survey findings and from notepads used while studying. According to the surveys, students for the most part studied using the system they were trained in, although some reported reverting back to preferred methods. However, many students in the no supplement groups reported that they did have sufficient study time to use their newly learned study systems. Meanwhile, none of the students who received supplements or used preferred methods reported that they lacked study time. Perhaps the study systems would have worked better had training-only students had more time to apply them. Several students across treatment groups also reported that they had to consciously resist reverting back to preferred methods, even if they reported that the new system (SQ3R or SOAR) was easy to use and easy to learn. This combination of findings indicates that students might simply need more training and practice using a new system before they are comfortable abandoning preferred methods in favor of the new system. In essence, old habits die hard.

The study notepads also confirmed that most students made sincere attempts to use their trained system (SQ3R or SOAR). However, although students in the SOAR/no supplement group were able to complete the Select and Organize steps proficiently, their notepads often lacked associations and regulation questions. These omissions might be due to lack of time or to lack of mastery of the SOAR system. Students in the SQ3R/no supplement group successfully used the first three steps of SQ3R (Survey, Question, Read), but it is difficult to assess if they followed through with the Recite and Review
steps. Some students reverted back to preferred methods altogether. It was especially troubling that students in both supplement groups spent time taking notes, even though they were given a set of complete study materials. Often their created notes followed the system they were trained in, so they essentially ended up with a product similar to what they were already given. Perhaps students believe that the act of creating materials helps them learn more than reading already created materials. If this is the case, then perhaps a SOAR or SQ3R framework that students complete (like that used in Jairam & Kiewra, 2010) would be more helpful to students than completed materials.

**Study Limitations and Future Research Directions**

The present study has several strengths. First, it is one of only two studies to compare SQ3R and SOAR head-to-head (Jairam, Kiewra, Rogers, Patterson-Hazley, & Marxhausen, 2012). Second, only the present study compares the two study systems in terms of training effects only. Third, the present study included concept training for the SOAR group, which is important because concept learning is part of the SOAR system. Fourth, the present study also used new content that involved skill learning. Thus, it was the first to examine the effects of SOAR and SQ3R on skill learning. Assessing skills is important in skill domains such as mathematics, science, and writing. Overall, the present study added several new dimensions and was more than a replication of the previous study. Because the present study is novel and has ventured into new realms of research, it should be considered a preliminary study. As such, there were some limitations that might have affected the results and that should be remedied in future research.
One possible limitation is that the amount of training and practice time provided for learning the new system (SQ3R or SOAR) was not sufficient for students to understand and apply the new system effectively. Although many students attempted to use their trained system, and even reported that it was easy to learn and use, several reverted back to preferred methods. Research has shown that the effectiveness of SQ3R increases when students are given more practice (Feldt, Byrne, & Bral, 2006). It is possible that the same is true of SOAR. Therefore, future research might provide additional training and practice using the new system prior to having students use it to learn new material.

Another possible limitation is that the amount of study time for the no supplement groups was insufficient. Those groups were asked to perform additional work (creating study materials) compared to supplement groups (who only had to study provided materials), but were not given additional time. Many students in the no supplement groups reported that they were rushed and did not have enough time to work through the entire system (SQ3R or SOAR). On the other hand, none of the students in supplement groups reported a lack of time. Consequently, the no supplement groups might have been at a disadvantage. Future research should take this discrepancy into account by perhaps providing additional time for students who create their own study materials.

**Instructional Implications and Conclusions**

The limitations and future research directions presented here offer hope for more fully discovering the relative benefits of SOAR and SQ3R training and supplements. In the meantime, recall that this thesis began by revealing a two-pronged problem: students
who do not know how to study and instructors who do not teach students how to study. SOAR has shown promise in past research as a good alternative to SQ3R and students’ preferred methods, and therefore its use should be recommended to instructors and students alike. Despite the limitations of the present study, SOAR merits further research to determine how it can best be applied to help students learn.

Understanding what study methods work best is a worthwhile research target warranting further investigation. Perhaps through increased understanding of SOAR training and application, educators can better help college students succeed in their courses, and, in a greater sense, throughout their lives.
REFERENCES


Appendix A: Demographic Information

1. Gender
   a. Male
   b. Female

2. Class Standing
   a. Freshman
   b. Sophomore
   c. Junior
   d. Senior

3. What is your overall GPA?
   a. 3.5 – 4.0
   b. 3.0 – 3.4
   c. 2.5 – 2.9
   d. 2.0 – 2.4
   e. Below 2.0

4. Have you ever studied the topic of Educational Measurement before?
   a. Yes
   b. No

5. Have you ever used the SQ3R study system?
   a. Yes
   b. No

6. Have you ever used the SOAR study system?
   a. Yes
   b. No
Appendix B: Educational Measurement Text
*Formatting adjusted from original

EDUCATIONAL MEASUREMENT

For those who construct and administer tests and interpret their results, an adequate understanding of measurement principles is essential. This text examines the fundamental measurement principles that underlie tests. It examines the three basic measures used to describe sets of test scores: central tendency, dispersion, and shape of distributions.

Measures of central tendency are indices of the typical or average level of performance and provide a reference point against which to compare other scores. Measures of dispersion reveal how clustered or spread out the scores are relative to their central tendency. The shape of the distribution is determined by the frequency of different scores in the distribution.

Measures of Central Tendency

Central tendency reveals the center for a set of scores. There are three basic ways of representing the central tendency of a set of test scores: the mode, the median, and the mean. Each represents a different way of defining the center of a set of scores.

The Mode

The mode is the score that occurs the most; that is, the score that is earned by the greatest number of individuals. Suppose five students had taken a quiz and earned the following scores: John, 10; Mary, 13; Joey, 8; Bobby, 7; and Susan, 7. Seven occurs twice in the distribution, while all other scores occur only once. Therefore, 7 is the mode. The mode is unaffected by extreme scores. For example, changing the highest score to 50 leaves the mode unaffected.

The Median

The median is the score that divides a distribution into two equal parts. It considers all the scores but not the size of the scores. The median is the point at which half of the remaining scores are higher and half are lower. To find the median, order the scores from smallest to largest and start counting from either end until you get to the middle. That middle-most score is the median. In the example presented in this section (test scores of 13, 10, 8, 7, and 7) the median score is 8; four scores remain, two higher than 8 and two lower. With an even number of scores, calculate the mean of the two middlemost scores to find the median. The median is unaffected by extreme scores. Note that if we change the highest score to 40 or 50, or something even higher, the median is unaffected.
The Mean

The average of a set of scores is called the **mean**. The mean is defined as the sum of all scores divided by the number of scores. To find the mean, simply add up all the scores and divide the total by the number of scores added. When the scores in our example (13, 10, 8, 7, and 7) are added together, the sum of scores is 45. Five scores were used, so we divide 45 by 5 and get a mean, or average score, of 9. The mean is the most often cited measure of central tendency. The mean can be unduly influenced by extreme scores because it considers the size of all scores. If John has $0.20 in his pocket; Mary has $0.35; Joey $0.15; Bobby $0.30; and Susan $120, then the average pocket money for the students would be $24.20. When there are extreme scores in the set, another measure of central tendency might better represent the typical score.

**Measures of Dispersion**

Although the mean, median, and mode are important indices in understanding a set of test scores, we also need to evaluate the dispersion of the scores. Measures of dispersion indicate the extent to which scores spread out or cluster around the center of the distribution. We examine three measures of dispersion: range, variance, and standard deviation.

**The Range**

The **range** is the simplest measure of dispersion. The range is simply the difference between the highest and lowest score in a set of scores. Therefore, finding the range is simply a matter of subtracting the lowest score from the highest. Returning to our sample test scores (13, 10, 8, 7, and 7), the range would be 13 – 7 = 6. The range is the simplest and easiest measure of dispersion because it depends exclusively on the value the highest and the lowest scores. The extent of spread or clustering of the other scores is ignored. To find measures of dispersion that characterize the dispersion of all scores, we must use more sophisticated metrics.

**Variance and Standard Deviation**

Measures of dispersion indicate the spread or clustering of scores about the center of the scores. Thus, the more direct approach to describing the dispersion of a set of scores would be to find the average distance of the scores from the center of the distribution. This is precisely what finding the **variance** and **standard deviation** is all about.

To describe the dispersion of the scores, one need only average the distance of each score from the mean. We call the distance between each score and the mean the **deviation score**. But, a set of scores balances at the mean because the deviation scores sum to zero. That is, the deviation scores with positive values (from scores that are larger than the mean) and those with negative values (from scores that are smaller than the mean) cancel...
each other out. For any set of scores, regardless of the number or size of the scores, the deviation scores always sum to zero. Since the average deviation score also is always zero (0 divided by any number is still 0), finding the average deviation score would not be very informative. Therefore, something else must be done.

The solution is to square each of the deviation scores. Since squaring a number always produces a positive result, whether the original number is negative or positive, squaring the deviation scores before adding them produces an informative result, known as the sum of squares. When this sum is divided by the number of scores, we have found the variance. The variance is the sum of the squared deviation scores divided by the number of scores in the distribution. Returning to our example, we find the sum of the squared deviation scores by subtracting each score from the mean (9 - 13 = -4, 9 - 10 = -1, 9 - 8 = 1, 9 - 7 = 2, 9 - 7 = 2), squaring each result (-4^2 = 16, -1^2 = 1, 1^2 = 1, 2^2 = 4, 2^2 = 4), summing the squares (16 + 1 + 1 + 4 + 4 = 26), and dividing the sum of squares by the number of scores (26/5 = 5.2).

Now, we have squared all the deviation scores to overcome the fact that the deviation scores themselves always add up to zero. But with the variance, we are left with the average of the squared deviation scores. That is, if we measured our students weight in pounds, their variance would be in squared pounds. If we measured their height in inches, the variance would be in square inches. Fortunately, this limitation is easily solved. We simply take the square root of the variance to get back to familiar units like pounds and inches. Using our previous example, we simply calculate the square root of the variance (√5.2) and find the standard deviation (2.28). The standard deviation is simply the square root of the variance.

It is important to note that the variance and standard deviation for any set of scores are perfectly redundant. That is, if you know the value of one, no additional information is needed to find the other. In statistics, it is sometimes convenient to refer to the variance rather than the standard deviation, but essentially they are two sides of the same coin.

**The Shape of the Distribution**

Although measures of central tendency and dispersion tell us much about a set of scores, sometimes we want to know the overall shape or pattern of a group of scores. Suppose that in a large chemistry class of 200 students, the distribution of scores provided below occurred on a quiz that contained 20 questions. Such a summary of scores is called a frequency distribution. A frequency distribution is a representation of the number of occurrences (i.e., frequency) of each score in a set of scores (i.e., the distribution).

Frequency distributions can be plotted on a graph where one axis represents the actual score values and the other axis represents the number of students who obtained that score. Figure 1 shows the graph of the frequency distribution for the scores reported below.
For a smaller classroom, say 43 students, we can quickly get a picture of the shape of the distribution by writing each different score on a separate line of notebook paper. When scores are the same, they are written side by side. For a 25-item classroom test with 43 students, our distribution might look like this:

<table>
<thead>
<tr>
<th>Score Obtained</th>
<th>Number of Students with that Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>1</td>
</tr>
<tr>
<td>18</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>4</td>
</tr>
<tr>
<td>16</td>
<td>9</td>
</tr>
<tr>
<td>15</td>
<td>10</td>
</tr>
<tr>
<td>14</td>
<td>19</td>
</tr>
<tr>
<td>13</td>
<td>27</td>
</tr>
<tr>
<td>12</td>
<td>42</td>
</tr>
<tr>
<td>11</td>
<td>27</td>
</tr>
<tr>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

Figure 1: Graph of the Frequency Distribution
By drawing a smooth line around the scores recorded in this fashion, we can get a rough look at the shape of our distribution. Draw a line around the scores in this example, and compare the resulting frequency curve to the one represented in Figure 1. What do these two frequency distributions have in common?

The major similarity is that both distributions are thicker in the middle. The most frequently occurring scores were those in the middle of the distribution. As you move from the middle to higher or lower scores, you find that fewer students obtained those scores. Also, moving away from the middle of the distribution, the observed frequency (i.e., number of students) falls off at about the same rate whether you are moving towards the higher scores or lower scores. A further similarity between these distributions is that both bear a family resemblance to what is called the normal distribution.

A normal distribution is a frequency distribution that corresponds closely to the distribution of scores that actually are obtained for many educational and psychological tests. A graph of this theoretical distribution is presented in Figure 2. A normal distribution is unimodal and symmetrical. That is, there is only one peak and the shape of the distribution on one side of that peak is the mirror image of the shape on the other side. In a normal distribution, the mean, median, and mode are always identical, making any of them a suitable measure of central tendency. The normal distribution is sometimes referred to as the normal curve or bell curve, because of its shape. In a normal distribution, most of the scores occur in the middle. In fact, about two-thirds of the scores fall in the region within one standard deviation on either side of the mean. About 95% of the scores fall within two standard deviations of the mean.

**Figure 2. Normal Distribution**
*The vertical line at 0 represents mean, median, and mode*
Distributions do not always resemble the normal curve. Sometimes they take on quite different shapes. Figure 3 compares the normal distribution to non-normal frequency distributions, called skewed distributions. A skewed distribution is an asymmetrical distribution in which scores tend to cluster at one end of the scale and taper off at the other.

**Figure 3. Skewed vs. Normal Distributions of Scores**

![Skewed vs. Normal Distributions of Scores](image)

Figure 3 shows both negatively and positively skewed distributions in relation to the normal distribution. Image (a) is **negatively skewed** because the narrow end or tail of the distribution points toward the low or negative end. When a distribution has a negative skew, the mode is greater than the median, and the median is larger than the mean. Under such circumstances, the median may be a better representative of the central tendency of the set of scores. Though the mean still lies at the center of the distribution, it does accurately reflect the majority of the scores. When students have mastered the course material well, teacher-made tests will result in a negatively skewed distribution of scores. Scores will cluster at the high end with a few scores tailing off to the low end. Such a distribution also can result if the test is just too easy. Examining a frequency distribution can reveal a great deal about the adequacy of learning and instruction.

Image (c) depicts a distribution of scores that is **positively skewed**. In a positively skewed distribution, the scores lump together at the lower end of the scale, causing the tail to point toward the higher or positive end. In this case, the mean is larger than the median, and the median is larger than the mode. Again, however, if the skew is great enough, you may wish to use the median as the main measure of central tendency, due to the mean being affected by extreme scores. People’s incomes typically are positively skewed, with most people making less than $40,000, but a few individuals pulling down incomes in the millions, tens of millions, and beyond.

Frequency distributions, like camels, sometimes have two humps. Figure 4 shows a **bimodal distribution**, a distribution with two modes or humps. Such a distribution might
be expected on a pop quiz on a day when only about half of the students have studied the assignment. If those students who studied did much better than those who did not, the separation between the scores of the two groups of students might make the distribution appear bimodal (See Figure 4). With a bimodal distribution, reporting a single mean or median will not accurately describe the situation, because this distribution is essentially a composite of two sets of scores, each with its own mode. In Figure 4, the mean and median would occur between the two humps, but would be a poor reflection of the center of the distribution. For example, if a company paid most workers $20,000, and most executives about $200,000, and had almost as many executives as workers, its salary structure would be bimodal, resembling the distribution pictured in Figure 4. In such a case, reporting only that the average (mean) income of all employees was $105,000 would misrepresent the nature of the distribution.

**Figure 4. A Bimodal Distribution**

![Figure 4](image)

Figure 4 represents the distribution that might occur when only half of the students are prepared for a pop quiz. It has two modes, with a large number of students scoring around 8/20 and a large number of students scoring around 16/20.
Appendix C: SQ3R Ideal Materials
*Formatting adjusted from original

SURVEY: *Skim text headings to get an overview of it.*
*Write a statement that predicts what the text is about.*

This text is about the three basic measures used to describe sets of test scores: central tendency, dispersion, and the shape of the distribution.

QUESTION: *Turn headings into written questions.*

What are the three basic measures used to describe sets of test scores?
What are the different measures of central tendency?
What is the mode?
What is the median?
What is the mean?
What are the different measures of dispersion?
What is the range?
What are variance and standard deviation?
What are the different shapes of distributions?
What is a frequency distribution?
What is a normal distribution?
What are skewed distributions?
What is a bimodal distribution?

READ: *Read the text and write answers to your questions.*

What are the three basic measures used to describe sets of test scores?

The three basic measures used to describe sets of scores are central tendency, dispersion, and the shape of the distribution.

What are the different measures of central tendency?

Central tendency reveals the center of a set of scores. The three measures of central tendency are the mean, median, and mode.

What is the mode?

The mode is the score that occurs the most – the score that is earned by the greatest number of individuals. It is unaffected by extreme scores because it is based only on the score that occurs most often.
What is the median?

The median is the middle-most score; the score that divides a distribution into two equal parts. In a set with an even number of scores, the mean of the two middle scores is the median. The median is unaffected by extreme scores because it considers all scores, but not their size.

What is the mean?

The mean is the average score – the sum of all scores divided by the number of scores. The mean is influenced by extreme scores because it considers the size of all scores.

What are the different measures of dispersion?

Dispersion indicates the extent to which scores spread out or cluster around the center of the distribution. The three measures of dispersion are range, variance, and standard deviation.

What is the range?

The range is the difference between the highest and lowest score in a set of scores. It is the simplest measure of dispersion. It does not give any information about the spread or clustering of the other scores in a set.

What are variance and standard deviation?

The variance and standard deviation both indicate the average distance of scores from the center of the distribution. The variance is defined as the sum of the squared deviation scores divided by the number of scores. It is calculated by finding how far each score is from the mean (deviation score), squaring those, adding them up (sum of squares), and dividing the sum of squares by the number of scores. The problem with variance is that you end up with squared deviation scores. In order to fix this, you take the square root of the variance, which gives you the standard deviation. If you have either the variance or the standard deviation, it is very simple to find the other, as the standard deviation is simply the square root of the variance. Both standard deviation and variance take into account the spread or clustering of all scores.

What are the different shapes of distributions?

The different shapes of a distribution are normal, skewed (positively or negatively), and bimodal.
What is a frequency distribution?

A frequency distribution is a representation of the number of occurrences of each score in a set of scores.

What is a normal distribution?

A normal distribution, also called a bell curve, is unimodal and symmetrical. It has one peak, and the shape of the distribution on one side of the peak is a mirror image of the shape on the other side. In a normal distribution, the mean, median, and mode are identical, and 95% of scores fall within two standard deviations of the mean.

What are skewed distributions?

Skewed distributions occur when scores cluster at one end of the scale and taper off at the other. Skewed distributions can be positively skewed or negatively skewed. In a negatively skewed distribution, scores cluster on the right side of the distribution and the tail points towards the left; the mode is greater than the median, and the median is greater than the mean. In a positively skewed distribution, scores cluster on the left side, and the tail points to the right; the mean is larger than the median, and the median is larger than the mode. The median is typically the best measure of central tendency with skewed distributions.

What is a bimodal distribution?

A bimodal distribution is a distribution with two modes, or humps. Reporting a single mean or median will not accurately describe a bimodal distribution.
RECITE: Recite your answers from memory. 
Do this first in your mind and then write them.

At this point, recite the answers to the following questions in your mind. You would then write them down, but since our goal is to provide you with ideal materials, we have done that second step for you on the following page.

What are the three basic measures used to describe sets of test scores?
What are the different measures of central tendency?
What is the mode?
What is the median?
What is the mean?
What are the different measures of dispersion?
What is the range?
What are variance and standard deviation?
What are the different shapes of distributions?
What is a frequency distribution?
What is a normal distribution?
What are skewed distributions?
What is a bimodal distribution?
What are the three basic measures used to describe sets of test scores?

The three basic measures used to describe sets of scores are central tendency, dispersion, and the shape of the distribution.

What are the different measures of central tendency?

Central tendency reveals the center of a set of scores. The three measures of central tendency are the mean, median, and mode.

What is the mode?

The mode is the score that occurs the most – the score that is earned by the greatest number of individuals. It is unaffected by extreme scores because it is based only on the score that occurs most often.

What is the median?

The median is the middle-most score; the score that divides a distribution into two equal parts. In a set with an even number of scores, the mean of the two middle scores is the median. The median is unaffected by extreme scores because it considers all scores, but not their size.

What is the mean?

The mean is the average score – the sum of all scores divided by the number of scores. The mean is influenced by extreme scores because it considers the size of all scores.

What are the different measures of dispersion?

Dispersion indicates the extent to which scores spread out or cluster around the center of the distribution. The three measures of dispersion are range, variance, and standard deviation.

What is the range?

The range is the difference between the highest and lowest score in a set of scores. It is the simplest measure of dispersion. It does not give any information about the spread or clustering of the other scores in a set.
What are variance and standard deviation?

The variance and standard deviation both indicate the average distance of scores from the center of the distribution. The variance is defined as the sum of the squared deviation scores divided by the number of scores. It is calculated by finding how far each score is from the mean (deviation score), squaring those, adding them up (sum of squares), and dividing the sum of squares by the number of scores. The problem with variance is that you end up with squared deviation scores. In order to fix this, you take the square root of the variance, which gives you the standard deviation. If you have either the variance or the standard deviation, it is very simple to find the other, as the standard deviation is simply the square root of the variance. Both standard deviation and variance take into account the spread or clustering of all scores.

What are the different shapes of distributions?

The different shapes of a distribution are normal, skewed (positively or negatively), and bimodal.

What is a frequency distribution?

A frequency distribution is a representation of the number of occurrences of each score in a set of scores.

What is a normal distribution?

A normal distribution, also called a bell curve, is unimodal and symmetrical. It has one peak, and the shape of the distribution on one side of the peak is a mirror image of the shape on the other side. In a normal distribution, the mean, median, and mode are identical, and 95% of scores fall within two standard deviations of the mean.

What are skewed distributions?

Skewed distributions occur when scores cluster at one end of the scale and taper off at the other. Skewed distributions can be positively skewed or negatively skewed. In a negatively skewed distribution, scores cluster on the right side of the distribution and the tail points towards the left; the mode is greater than the median, and the median is greater than the mean. In a positively skewed distribution, scores cluster on the left side, and the tail points to the right; the mean is larger than the median, and the median is larger than the mode. The median is typically the best measure of central tendency with skewed distributions.
What is a bimodal distribution?

A bimodal distribution is a distribution with two modes, or humps. Reporting a single mean or median will not accurately describe a bimodal distribution.

**REVIEW: Reflect on your questions and answers. Consider if the questions are appropriate and if your answers are correct.**

*We believe these questions are effective and appropriate and that the answers are correct and thorough, with all relevant information included.*
There are three main basic measures used to describe sets of scores obtained from tests. These three measures are central tendency, dispersion, and shape of the distribution.

**Three Measures of Central Tendency**

Central tendency reveals the center of a set of scores. There are three different ways to define the center of a set of scores: mean, median, and mode.

**Mean**
- The average score
- Found by dividing the sum of all scores by the number of scores
- Considers the size of all scores – therefore is affected by extreme scores in the set

**Median:**
- The middle-most score – the score that divides a distribution into two equal parts
- When there is an even number of scores in a set, find the median by calculating the mean of the two middlemost scores
- Considers all scores but not the size of the score – therefore is not affected by extreme scores

**Mode:**
- The score that occurs most often
- Is only based on a few scores in the set – therefore is not affected by extreme scores

**Three Measures of Dispersion**

Dispersion indicates the extent to which scores spread out or cluster around the center of the distribution. There are three measures of dispersion: range, variance, and standard deviation.

**Range:**
- The difference between the highest and lowest score in a set of scores
- Simplest measure of dispersion
- Depends exclusively on highest and lowest scores – extent of spread or clustering of other scores is ignored
Variance:
- Indicates the average distance of scores from the center of the distribution.
- Defined as the sum of the squared deviation scores divided by the number of scores
  - Calculated by finding how far each score is from the mean (deviation score), squaring these, adding them up (sum of squares), and dividing the sum of squares by the number of scores
- Is more complex than range – takes into account the spread or clustering of all scores
- Limitation is that you end up with squared deviation scores

Standard Deviation:
- Like variance, standard deviation indicates the average distance of scores from the center of the distribution.
- Calculated by finding the square root of the variance
- This solves the problem of having squared deviation scores
- Like variance, takes into account the spread or clustering of all scores

Three Shapes of Distributions: Normal, Skewed, and Bimodal

The shape of pattern created by graphing a group of scores can give us information about the set of scores. First, create a frequency distribution (a representation of the number of occurrences of each score in a set of score). Then plot the frequency distribution on a graph to reveal the shape of the distribution.

Normal:
- Unimodal and symmetrical (It has one peak, and the shape of the distribution on one side of the peak is a mirror image of the shape on the other side)
- Also called the bell curve due to its shape
- The mean, median, and mode are identical.
- In a normal distribution, 95% of scores fall within 2 standard deviations of the mean.

Skewed:
- Occur when scores cluster at one end of the scale and taper off at the other
- Can be skewed either positively or negatively
- In a negatively skewed distribution, scores cluster on the right side and the tail points toward the left
  - The mode is greater than the median, and the median is greater than the mean
- In a positively skewed distribution, scores cluster on the left side and the tail points toward the right
The mean is larger than the median, and the median is larger than the mode.
- The median is typically the best measure of central tendency with skewed distributions.

Bimodal:
- A distribution with two modes, or humps
- Reporting a single mean or median will not accurately describe a bimodal distribution

ORGANIZE

Educational Measurement

- Central Tendency
  - Mean
  - Median
  - Mode

- Dispersion
  - Range
  - Variance
  - Standard Deviation

- Shape of Distribution
  - Normal
  - Skewed
    - Positively
    - Negatively
  - Bimodal
### Central Tendency

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Median</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Definition:</strong></td>
<td>Sum of all scores divided by the number of scores</td>
<td>Score that divides a distribution into 2 equal parts</td>
<td>Score that occurs most often</td>
</tr>
<tr>
<td><strong>Sample scores:</strong></td>
<td>Set of scores = 10, 13, 8, 7, 7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| **Calculation:** | 10  
|                | 13  
|                | 8   | Order scores from smallest to largest: 7, 7, 8, 10, 13  
|                | 7   | *8 is middle-most score  
|                | +7  | 7 occurs most often |                                |
|                | 45  |                                |                                |
|                | 45/5 = 9 |                                |                                |
| **Extreme Scores:** | Do have effect | Have no effect | Have no effect |
| **Complexity:** | Considers the size of all scores | Considers all scores, but not their size | Based upon a few scores |

### Dispersion

<table>
<thead>
<tr>
<th></th>
<th>Range</th>
<th>Variance</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Definition:</strong></td>
<td>Difference between highest and lowest scores</td>
<td>Sum of the squared deviation scores divided by the number of scores</td>
<td>Square root of the variance</td>
</tr>
<tr>
<td><strong>Sample scores:</strong></td>
<td>13, 10, 8, 7, 7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| **Calculation:** | 13 – 7 = 6 | Mean – Score = Dev. Score 9 – 13 = -4  
|                |             | 9 – 10 = -1  
|                |             | 9 – 8 = 1  
|                |             | 9 – 7 = 2  
|                |             | 9 – 7 = 2  
|                |             | -4² + -1² + 1² + 2² + 2² becomes 16 + 1 + 1 + 4 + 4 = 26  
<p>|                |             | 26/5 = 5.2 |
| <strong>Complexity:</strong> | Simplest. Depends on only two scores | Indexes the spread or clustering of all scores | Indexes the spread or clustering of all scores |
| <strong>Limitation:</strong> | Ignores spread or clustering of other scores | Deviation scores are squared | None |</p>
<table>
<thead>
<tr>
<th>Shape of the Distribution</th>
<th>Description</th>
<th>Illustration:</th>
<th>Best Measure of Central Tendency:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bimodal</td>
<td>Two modes (two humps)</td>
<td><img src="image" alt="Bimodal Illustration" /></td>
<td>Both modes</td>
</tr>
<tr>
<td>Positively Skewed</td>
<td>Scores cluster on the right - tail points to the right</td>
<td><img src="image" alt="Positively Skewed Illustration" /></td>
<td>Median</td>
</tr>
<tr>
<td>Negatively Skewed</td>
<td>Scores cluster on the left - tail points to the left</td>
<td><img src="image" alt="Negatively Skewed Illustration" /></td>
<td>Median</td>
</tr>
<tr>
<td>Normal</td>
<td>Unimodal and symmetrical</td>
<td><img src="image" alt="Normal Illustration" /></td>
<td>Any, they are the same</td>
</tr>
</tbody>
</table>

For the unimodal and symmetrical distribution, the measures of central tendency are: a = mean, b = median, c = mode.
ASSOCIATE

Associations Within the Material
1. The number 3 comes up a lot: three basic measures of sets of scores; three measures of central tendency; three measures of dispersion; and three shapes of the distribution.
2. The mean is the only measure of central tendency affected by extreme scores, whereas mean and median are not.
3. Range is the least complex measure of dispersion because it only takes the largest and smallest scores into account.
4. Standard deviation is simply the square root of the variance; therefore, if you have calculated the variance, it is very easy to determine the standard deviation.
5. A frequency distribution is a way to organize scores, and can be used to create a graph of the distribution.
6. In a positively skewed distribution, the tail points at the positive (right) end, whereas in a negatively skewed distribution, the tail points at the negative (left) end.
7. The median is the best indicator of central tendency in a skewed distribution, whereas both modes should be reported to best describe the central tendency of a bimodal distribution. Any measure of central tendency would be appropriate to describe a normal distribution.
8. The normal curve is also called the bell curve because its shape resembles a bell.

Associations Outside the Material
1. A median is what divides a highway in half – the median is the measure of central tendency that divides a set of scores in half.
2. “Mode” sounds similar to “most” – the mode is the score that occurs most often.
3. A cheetah’s range is the distance from one end of their territory to the other – in a set of scores, the range is the distance between the highest and lowest score. Essentially, the distance from one end of the set to the other.
4. To deviate means to stray from the average/norm – standard deviation is a way to show the average distance from the average score.
5. If the average commute time of the people in your office is 23 minutes, but you want to know more about the dispersion of commute times, it makes sense to use standard deviation, because reporting variance would give you the answer in squared minutes.
6. A simple first-letter mnemonic to help you remember the four types of distributions is “Nelly Nibbles Peanut Butter” (Normal, Negatively skewed, Positively skewed, Bimodal).
7. “Bi” means two; therefore, bimodal means “two modes”
Fact Questions:
1. What are the three basic measures used to describe sets of scores?
2. What are the three measures of central tendency?
3. What score divides a set of scores?
4. What is the score that occurs the most called?
5. What do you find if you calculate the average score?
6. What are the different shapes a distribution can have?
7. In a normal distribution, where do most of the scores fall?
8. In a negatively skewed distribution, where does the mean fall?
9. What is the name of a distribution with two humps?
10. What are the three measures of dispersion?
11. What measure of dispersion do you calculate by squaring the standard deviation?
12. How do you calculate the range?
13. What is a frequency distribution?

Relationship Questions:
1. In which distribution(s) is the mode greater than the mean?
2. Which measure(s) of central tendency is/are unaffected by extreme scores?
3. What is the difference between variance and standard deviation?
4. Which type(s) of distribution(s) has/have many low scores and a few high scores?
5. Which type(s) of distribution(s) has/have a bell shaped curve?

Concept Questions:
1. The high temperature on Tuesday was 68; the low was 32. What can we calculate from this information?
2. Looking over the height chart for her volleyball team, the coach said, “It looks like most of you are 5’8’’. What concept has the coach expressed?
3. Mrs. O’Leary’s students received scores of 72, 77, 83, 89, and 94. Mrs. O’Leary did the following: 72+77+83+89+94 = 415/5 = 83. What did she find?
4. In math class, most students got A’s on the last pop quiz. A couple of students missed class the day before and got D’s. What type of distribution will these scores create?
5. In a classroom, all of the students are between 5’4’’ and 5’10’’ tall. A new student, Alex, is 6’5’’ tall. In a distribution of this information, which measure of central tendency would be affected by the addition of Alex?
6. Peter finds out the average weight of sumo wrestlers is 350 pounds, and that most wrestlers typically fall within 50 pounds of that weight. This typical distance from the mean represents what?

(Try to answer these on your own first. Answers can be found on the following page)
Answers to Regulation Questions:

Fact Questions:
1. Central tendency, dispersion, and the shape of the distribution
2. Mean, median, and mode
3. Median
4. Mode
5. Mean
6. Normal, skewed (positively or negatively), bimodal
7. In the center, 95% fall between 2 standard deviations of the mean
8. On the left side of the graph – to the left of both median and mode
9. Bimodal distribution
10. Range, variance, and standard deviation
11. Variance
12. Find the difference between the highest and lowest scores
13. A representation of the number of occurrences of each score in a set of scores.

Relationship Questions:
1. Negatively skewed and bimodal (in bimodal, one mode is greater than the mean)
2. Median and mode are unaffected by extreme scores
3. Standard deviation is simply the square root of the variance
4. Positively skewed distribution
5. The normal distribution

Concept Questions:
1. Range
2. Mode
3. Mean
4. Negatively skewed distribution
5. Mean
6. Standard deviation
Appendix E: Vocabulary Quiz

Vocabulary Questions: Circle the Correct Letter

1. Today, Wegener’s theory is _____; however, he died an outsider treated with _____ by the scientific establishment.
   
   A. unsupported; approval
   B. dismissed; contempt
   C. accepted; approbation
   D. unchallenged; disdain
   E. unrivalled; reverence

2. The revolution in art has not lost its steam; it ____ on as fiercely as ever.

   A. trudges
   B. meanders
   C. edges
   D. ambles
   E. rages

3. Each occupation has its own ____; bankers, lawyers, and computer professionals, for example, all use among themselves language which outsiders have difficulty following.

   A. merits
   B. disadvantages
   C. rewards
   D. jargon
   E. problems

4. ____ by nature, Jones spoke very little even to his own family members.

   A. garrulous
   B. equivocal
   C. taciturn
   D. arrogant
   E. gregarious
5. Biological clocks are of such ____ adaptive value to living organisms, that we would expect most organisms to ____ them.

A. clear; avoid
B. meager; evolve
C. significant; eschew
D. obvious; possess
E. ambivalent; develop

6. The peasants were the least ____ of all people, bound by tradition and ____ by superstitions.

A. free; fettered
B. enfranchised; rejected
C. enthralled; tied
D. pinioned; limited
E. conventional; encumbered
Appendix F: Fact Achievement Test

1. Which of the following is not one of the basic measures used to describe a set of scores?
   a. Central tendency
   b. Dispersion
   c. Frequency distribution
   d. Shape of the distribution

2. Which measure of central tendency represents the score that occurs the most?
   a. Mean
   b. Median
   c. Mode
   d. Variance

3. What is the sum of all scores divided by the number of scores?
   a. Mean
   b. Median
   c. Mode
   d. Variance

4. How do you find the median of a set of 8 scores?
   a. Use the 4th score
   b. Use the 5th score
   c. Use the mean of the 2 middlemost scores
   d. Use the mean of the 4 middlemost scores

5. What is the best measure of central tendency for a negatively skewed distribution?
   a. Mean
   b. Median
   c. Mode
   d. Variance

6. In a positively skewed distribution, where do most of the scores fall?
   a. The lower end (left side) of the distribution
   b. The middle of the distribution
   c. The upper end (right side) of the distribution
   d. Equally distributed throughout the lower end, middle, and upper end
7. In which distribution is the mode always greater than the mean?
   a. Bimodal
   b. Negatively skewed
   c. Normal
   d. Positively skewed

8. In which distribution are the mean, median, and mode equivalent?
   a. Bimodal
   b. Negatively skewed
   c. Normal
   d. Positively skewed

9. What is the sum of the squared deviations divided by the number of scores?
   a. Frequency distribution
   b. Range
   c. Standard deviation
   d. Variance

10. What measure is the difference between the lowest and highest scores?
    a. Range
    b. Sum of squares
    c. Standard deviation
    d. Variance

11. Which measure of dispersion fails to consider how scores differ from the mean?
    a. Frequency distribution
    b. Range
    c. Standard deviation
    d. Variance

12. The standard deviation is the square root of which measure of dispersion?
    a. Frequency distribution
    b. Range
    c. Sum of squares
    d. Variance

13. A __________ is a representation of the number of occurrences of each score in a set of scores.
    a. Frequency distribution
    b. Negatively skewed curve
    c. Normal curve
    d. Positively skewed curve
14. In a normal distribution, approximately what percentage of scores fall within 2 standard deviations of the mean?
   a. 34
   b. 68
   c. 82
   d. 95

15. A frequency distribution that results in a graph with two humps is said to be
   a. Bimodal
   b. Normal
   c. Negatively skewed
   d. Positively skewed
Appendix G: Relationship Achievement Test

1. Which two types of central tendency are least affected by extreme scores?
   a. Mean and Median
   b. Mean and Mode
   c. Median and Mode
   d. Median and Range

2. Which two types of dispersion indicate the average distance of scores from the center of the distribution?
   a. Range and standard deviation
   b. Range and variance
   c. Standard deviation and variance
   d. Standard deviation and the sum of scores

3. How are positively and negatively skewed distributions alike?
   a. They both point to the right
   b. They both point to the left
   c. The mean is always larger than the mode
   d. The median is always between the mode and mean

4. Rank order the mean, median, and mode with respect to complexity, starting with the least complex.
   a. Mean, median, mode
   b. Median, mean, mode
   c. Mode, mean, median
   d. Mode, median, mean

5. Which measure of dispersion is least complex?
   a. Frequency distribution
   b. Range
   c. Standard deviation
   d. Variance
Appendix H: Concept Achievement Test

1. Mrs. Griffin found that the test scores of her students centered on the score of 87. She later found out that one of the test scores had to be changed from 68 to 0 because the student had cheated. What measure of central tendency would be most affected?
   a. Mean
   b. Median
   c. Mode
   d. Variance

2. There are 25 students in Ms. Brown’s class. She arranged their test scores from highest to lowest. The 13th score represents what?
   a. Mean
   b. Median
   c. Mode
   d. Variance

3. Students in Mr. Marsh’s class took a quiz and earned scores of 3, 5, 8, 7, 7, and 6. Mr. Marsh then did the following: $3 + 5 + 6 + 7 + 7 + 8 = 36/6 = 6$. What did he find?
   a. Mean
   b. Median
   c. Mode
   d. Variance

4. In terms of natural hair color, there are more brunettes than blondes or redheads. What measure of central tendency is illustrated here?
   a. Mean
   b. Median
   c. Mode
   d. Variance

5. The tallest boy on the team is 6’4”; the shortest boy is 5’0”. This information is sufficient for finding the:
   a. Median
   b. Range
   c. Standard deviation
   d. Variance
6. Sasha is doing research and found out the average age of chess masters is 24, and that most masters typically fall within 2 years of that age. This typical distance from the mean represents what?
   a. Range
   b. Sum of squares
   c. Standard deviation
   d. Variance

7. Most of Mrs. Cane’s students received Cs on the test. A lesser number had Bs and Ds. Very few had As and Fs. What type of distribution will these scores create?
   a. Bimodal
   b. Negatively skewed
   c. Normal
   d. Positively skewed

8. On a test of math achievement, students who had a course in fractions tended to get about 95% of the questions correct on the test. An equal number of students who had not had the course on fractions tended to get about 10% of the questions correct. What type of distribution will these scores create?
   a. Bimodal
   b. Negatively skewed
   c. Normal
   d. Positively skewed

9. In second grade most students are around 48 inches tall. A few students are over 60 inches tall. What type of distribution will these scores create?
   a. Bimodal
   b. Negatively skewed
   c. Normal
   d. Positively skewed

10. What type of distribution is represented in the picture below?

   ![Distribution Graph]

   a. Bimodal
   b. Negatively skewed
   c. Normal
   d. Positively skewed
Appendix I: Skill Achievement Test

For items 1-6, use the following set of scores: 2, 2, 4, 5, 7, 7, 8, 8, 8, 9

1. What is the mean?
   a. 5
   b. 6
   c. 7
   d. 8

2. What is the median?
   a. 5
   b. 6
   c. 7
   d. 8

3. What is the mode?
   a. 5
   b. 6
   c. 7
   d. 8

4. What is the range?
   a. 5
   b. 6
   c. 7
   d. 8

5. What is the variance?
   a. 5
   b. 6
   c. 7
   d. 8

6. What shape of distribution would this data set create?
   a. Bimodal
   b. Negatively skewed
   c. Normal
   d. Positively skewed
Appendix J: Attitude Surveys

Survey (Control)

1. The training session prepared me for studying the educational measurement material.
   a. Strongly Disagree
   b. Disagree
   c. Agree
   d. Strongly Agree

2. In the space below, describe how you studied the educational measurement material.
   Please be specific. Continue your answer on the back if you need to.
Survey (SQ3R and SOAR)

1. The training session prepared me for studying the educational measurement material.
   a. Strongly Disagree
   b. Disagree
   c. Agree
   d. Strongly Agree

2. The study system shown to me was easy to learn?
   a. Strongly Disagree
   b. Disagree
   c. Agree
   d. Strongly Agree

3. The study system shown to me was easy to use?
   a. Strongly Disagree
   b. Disagree
   c. Agree
   d. Strongly Agree

4. The study system shown to me was effective?
   a. Strongly Disagree
   b. Disagree
   c. Agree
   d. Strongly Agree

5. The study system shown to me was enjoyable to use?
   a. Strongly Disagree
   b. Disagree
   c. Agree
   d. Strongly Agree

6. The study system shown to me is one that I will reuse?
   a. Strongly Disagree
   b. Disagree
   c. Agree
   d. Strongly Agree

7. In the space below, describe how you studied the educational measurement material. Please be specific. Continue your answer on the back if you need to.