



การเปรียบเทียบผลที่ได้จากการทวนคำตอบในข้อสอบแบบหลายตัวเลือกที่มีการเรียงลำดับข้อสอบ
ระยะเวลาในการสอบ และระดับความสามารถของผู้สอบที่แตกต่างกัน

A COMPARISON OF CONSEQUENCES FROM REVIEWING ANSWERS IN MULTIPLE-CHOICE
ITEMS AMONG DIFFERENT TEST ITEM ORDER, RESPONSE TIME AND
EXAMINEES' ABILITY LEVELS

นางสาวจุฑาภรณ์ มาสันเทียะ *

Juthaporn Masantiah

รศ.ดร.โชติกา ภาชีผล **

Assoc. Prof. Shotiga Pasiphol, Ph.D.

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อ 1) ศึกษาความสัมพันธ์ระหว่างผลที่ได้จากการทวนคำตอบกับระดับความยากของข้อสอบและระหว่างผลที่ได้จากการทวนคำตอบกับระดับความสามารถของผู้สอบ 2) เปรียบเทียบผลที่ได้จากการทวนคำตอบที่มีการเรียงลำดับข้อสอบ ระยะเวลาในการสอบ และระดับความสามารถของผู้สอบที่แตกต่างกัน กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 4 ภาคเรียนที่ 2 ปีการศึกษา 2556 จากโรงเรียนในเขตพื้นที่การศึกษากรุงเทพมหานคร เขต 2 จำนวน 5 โรงเรียน คิดเป็นนักเรียน 338 คน ผลการวิจัยพบว่า 1) สัมประสิทธิ์สหสัมพันธ์ระหว่างผลที่ได้จากการทวนคำตอบกับระดับความยากของข้อสอบมีค่า $-.987 - .344$ 2) สัมประสิทธิ์สหสัมพันธ์ระหว่างผลที่ได้จากการทวนคำตอบกับระดับความสามารถของผู้สอบมีค่า $-.939 - .424$ 3) ค่าเฉลี่ยจำนวนข้อที่ทวนคำตอบแล้วยังคงคำตอบถูกไว้ (RR) ต่างกันในระยะเวลาสอบที่ต่างกัน อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และต่างกันในแต่ละระดับความสามารถของผู้สอบอย่างมีนัยสำคัญทางสถิติที่ระดับ.05 4) ค่าเฉลี่ยจำนวนข้อที่ทวนคำตอบแล้วยังคงคำตอบผิดไว้ (WW) ต่างกันในระยะเวลาสอบที่ต่างกัน อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และต่างกันในแต่ละระดับความสามารถของผู้สอบที่อย่างมีนัยสำคัญทางสถิติที่ระดับ.05 5) ค่าเฉลี่ยของจำนวนข้อที่ทวนคำตอบแล้วมีการเปลี่ยนคำตอบจากถูกเป็นผิด (R^*W) ต่างกัน เมื่อระยะเวลาในการสอบ และระดับความสามารถของผู้สอบแตกต่างกัน อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 6) ค่าเฉลี่ยของจำนวนข้อที่ทวนคำตอบแล้วเปลี่ยนจากผิดเป็นถูก (W^*R) ต่างกัน เมื่อระยะเวลาในการสอบ และระดับความสามารถของผู้สอบต่างกัน อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 7) ค่าเฉลี่ยของจำนวนข้อที่ทวนคำตอบแล้วเปลี่ยนจากผิดเป็นผิด (WW) ไม่แตกต่างกันเมื่อ มีการเรียงลำดับข้อสอบ ระยะเวลาในการสอบและระดับความสามารถของผู้สอบที่แตกต่างกัน

* นิสิตมหาบัณฑิตสาขาวิชาการศึกษาคณิตศาสตร์ ภาควิชาวิจัยและจิตวิทยาการศึกษา

คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

E-mail Address: star_cu55@hotmail.com

** อาจารย์ประจำสาขาวิชาการศึกษาวัดและประเมินผลการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา

คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

E-mail Address: Aimornj@hotmail.com

ISSN 1905-4491

Abstract

The purposes of this research were 1) to study the correlation between the consequences of reviewing answers and item difficulty and the correlation between consequences of reviewing answers and examinees' ability levels, and 2) to compare the consequences of reviewing answers in multiple-choice items among different test item order, response time and examinees' ability levels. The participants were 338 tenth grade students in Bangkok. The research findings were as follows: 1) There was a correlation (-.987 - .344) between consequences of reviewing answers and item difficulty; 2) there was a correlation (-.939 - .424) between consequences of reviewing answers and examinees' ability levels; 3) there were significant consequences of reviewing answers (RR) under different response time and under different examinees' ability levels at .05 level; 4) there were significant consequences from reviewing answers (WW) under different response time and under different examinees' ability levels at .05 level; 5) there were significant consequences from reviewing answers (R*W) among different response time and examinees' ability levels at .05 level; 6) there were significant consequences from reviewing answers (W*R) among different response time and examinees' ability levels at .05 level; and 7) there were no significant consequences from reviewing answers (W*W) among different response time and examinees' ability levels at .05 level on.

คำสำคัญ : การทวนคำตอบ / ข้อสอบแบบหลายตัวเลือก / การเรียงลำดับข้อสอบ / ระยะเวลาในการสอบ / ระดับความสามารถของผู้สอบ

KEYWORDS : REVIEWING ANSWERS / MULTIPLE-CHOICE ITEMS / TEST ITEM ORDER / RESPONSE TIME / EXAMINEES' ABILITY LEVELS

บทนำ

เครื่องมือการวัดและประเมินผลสัมฤทธิ์ทางการเรียนที่ครูผู้สอนหรือผู้สร้างข้อสอบนิยมเลือกใช้ในปัจจุบัน มักจะเป็นข้อสอบประเภทเลือกตอบมากกว่าข้อสอบประเภทเสนอคำตอบ (Kastner & Stangl, 2011; ประณาท เทียนศรี, 2546) และข้อสอบประเภทเลือกตอบที่เป็นข้อสอบหลายตัวเลือก (Multiple-Choice) เป็นที่นิยมใช้มากที่สุดในปัจจุบัน (Al-Hamly & Coombe, 2005) ด้วยคุณสมบัติที่วัดเนื้อหาได้ครอบคลุม ให้คะแนนอย่างเป็นปรนัย ประหยัดเวลาและค่าใช้จ่ายในการตรวจ สามารถใช้ในการสอบขนาดใหญ่ได้ และวัดในระดับพฤติกรรมขั้นสูงได้ (Haladyna & Downing, 1989; โชติกา ภาชีผล, 2555) ซึ่งในการบริหารและการจัดการในการสอบมีหลักปฏิบัติที่สำคัญ เช่น ให้เวลาสอบที่เพียงพอ เตรียมเอกสารในการสอบให้เพียงพอตลอดจน จัดเตรียมสถานที่ให้เหมาะสมปราศจากสิ่งรบกวน และควบคุมมาตรฐานในการคุมสอบของผู้คุมสอบ เป็นต้น ซึ่งจากผลงานวิจัยส่วนใหญ่พบว่า ครูผู้สอน หรือ ผู้คุมสอบควรกระตุ้นหรือแนะนำให้ผู้สอบทวนคำตอบหลังจากทำข้อสอบเสร็จแล้วเพราะ ผลที่ได้จากการทวนคำตอบนั้นจะส่งผลให้นักเรียนเปลี่ยนคำตอบและมีคะแนนเพิ่มขึ้น (Al-Hamly & Coombe, 2005; Bridgeman, 2012; Di Milia, 2007) ซึ่งรูปแบบของการเปลี่ยนคำตอบที่พบมี 3 รูปแบบ คือ 1) เปลี่ยนจากผิดเป็นถูก (Wrong-Right / Incorrect-Correct) 2) เปลี่ยนจากถูกเป็นผิด (Right-Wrong / Correct-Incorrect) 3) เปลี่ยนจากผิดเป็นผิด (Wrong-Wrong / Incorrect-Incorrect)

ในการศึกษาวิจัยที่เกี่ยวข้องกับผลที่ได้จากการทวนคำตอบในข้อสอบหลายตัวเลือกในระยะเวลา 40 ปีผ่านมามีแนวโน้มไปในทิศทางเดียวกัน คือ ครู หรือ ผู้คุมสอบควรแนะนำให้นักเรียนทวนคำตอบและเปลี่ยนคำตอบ (Al-Hamly & Coombe, 2005; Ferguson, Kreiter, Peterson, Rowat, & Elliott, 2002; Mueller & Wasser, 1977; กานดา ทองวัฒนะ, 2521; ประณาท เทียนศรี, 2546; พิมพ์สิริ เจริญนเรศธรรม,

2549) แต่ก็ยังมีบางผลงานวิจัยที่ได้กล่าวว่า ไม่ควรแนะนำให้ผู้สอบเปลี่ยนคำตอบถ้ามีโอกาสในการทวนคำตอบเนื่องจากจะทำให้ได้ผลเสีย (Foote & Belinky, 1972; Van der Linden, Jeon, & Ferrara, 2011) จึงเป็นปัญหาว่า การกระตุ้นให้นักเรียนทวนคำตอบแล้วเปลี่ยนคำตอบนั้นจะเป็นผลดีต่อผู้สอบจริงหรือไม่ เพราะในการศึกษาวิจัยเรื่องผลที่ได้จากการทวนคำตอบในแต่ละงานนั้นได้ทำการศึกษาในบริบทที่แตกต่างกัน และมีวิธีในการคิดคำนวณที่ต่างกัน แต่อย่างไรก็ดีจากการศึกษางานวิจัยที่ผ่านมา ทำให้เห็นว่าการเปลี่ยนคำตอบนั้นน่าจะส่งผลต่อคะแนนหรือความสามารถของผู้สอบ โดยปัจจัยส่วนใหญ่ที่จะทำให้เกิดการทวนคำตอบแล้วเปลี่ยนคำตอบนั้น ได้แก่ 1) เพศ โดยพบว่าเพศชายมีการเปลี่ยนคำตอบแล้วได้คะแนนสูงกว่าเพศหญิง (Di Milia, 2007; Reile & Briggs, 1952; ประณาท เทียนศรี, 2546) แต่ก็มีงานวิจัยบางงานที่พบว่าเพศหญิงมีการเปลี่ยนคำตอบแล้วได้คะแนนสูงกว่าเพศชาย Bath (1967 cited in Al-Hamly & Coombe, 2005) หรือในบางงานวิจัยพบว่า ทั้งเพศชายและเพศหญิงมีการเปลี่ยนคำตอบไม่แตกต่างกัน (ขวัญใจ อุดราภรณ์, 2528) 2) ระดับความยากของข้อสอบ ที่พบว่ามีความสัมพันธ์กับการเปลี่ยนคำตอบ (Al-Hamly & Coombe, 2005; Ferguson, Kreiter, Peterson, Rowat, & Elliott, 2002; Mueller & Wasser, 1977; กานดา ทองวัฒนะ, 2521) 3) ค่าอำนาจจำแนก ที่พบว่ามีความสัมพันธ์กับการเปลี่ยนคำตอบระหว่างการสอบ (Ferguson, Kreiter, Peterson, Rowat, & Elliott, 2002; กานดา ทองวัฒนะ, 2521; ขวัญใจ อุดราภรณ์, 2528) 4) ระดับความสามารถของผู้สอบที่พบว่า นักเรียนกลุ่มสูงมักจะเปลี่ยนคำตอบแล้วส่งผลให้ได้คะแนนสูงขึ้น (Di Milia, 2007; Ferguson, Kreiter, Peterson, Rowat, & Elliott, 2002; Vispoel, 1998; ประณาท เทียนศรี, 2546; พิมพ์สิริ เจริญนเรศวร, 2549) 5) ความวิตกกังวลในการสอบ (Mueller & Wasser, 1977) ที่พบว่า นักเรียนที่มีความวิตกกังวลต่อการสอบจะมีการเปลี่ยนคำตอบมากกว่านักเรียนที่ไม่มีความวิตกกังวล

จากปัจจัยที่ส่งผลให้เกิดการทวนแล้วเปลี่ยนคำตอบนี้ จะพบว่า มีนักวิจัยศึกษามาอย่างยาวนานและได้ผลที่ค่อนข้างชัดเจนแล้ว ผู้วิจัยจึงได้สนใจศึกษาปัจจัยที่เกี่ยวข้องกับคุณลักษณะของข้อสอบเพิ่มเติม ได้แก่ การเรียงลำดับข้อสอบ ซึ่งเทคนิคที่ครูหรือผู้สร้างข้อสอบใช้ในการป้องกันการลอกข้อสอบคือ การเพิ่มจำนวนชุดของข้อสอบ โดยภายในข้อสอบแต่ละชุดจะมีลำดับข้อสอบที่สลับกัน ซึ่งจะช่วยป้องกันการลอกข้อสอบหรือการโกงกันได้ แต่การนำข้อสอบมาสลับลำดับข้อคำถามนั้น ครูหรือผู้สร้างข้อสอบอาจไม่ได้คำนึงถึงลำดับความยากง่ายของข้อสอบ หรือลำดับของเนื้อหา ซึ่งส่งผลต่อความสามารถที่แท้จริงของผู้สอบได้ (ศิริวิมล สุขสมบูรณ์, 2541) โดยจากงานวิจัยที่เกี่ยวข้องกับการเรียงลำดับข้อสอบนั้นพบว่า การเรียงลำดับข้อสอบที่เหมาะสมและสะท้อนความสามารถที่แท้จริงของผู้สอบได้ดีที่สุด คือ การเรียงลำดับข้อสอบจากข้อง่ายไปยังข้อยากตามเนื้อหาของบทเรียน (Aamodt & McShane, 1992; Ahmann & Glock, 1971; ศิริวิมล สุขสมบูรณ์, 2541) ผู้วิจัยจึงมีความสนใจที่จะศึกษาการเรียงข้อสอบมีผลต่อการเปลี่ยนคำตอบหรือไม่อย่างไร

นอกจากการเรียงลำดับข้อสอบที่น่าจะส่งผลต่อการเปลี่ยนคำตอบแล้ว ปัจจัยที่น่าจะส่งผลต่อการเปลี่ยนคำตอบ คือ ระยะเวลาที่ใช้ในการจัดการสอบ โดย พิมพ์สิริ เจริญนเรศวร (2549) พบว่าผู้สอบส่วนใหญ่จะใช้เวลาในการทำข้อสอบในเงื่อนไขการให้ทวนคำตอบมากกว่าการไม่ให้ทวนคำตอบเฉลี่ย 19.90% ซึ่งผู้สอบที่มีระดับความสามารถแตกต่างกันก็จะใช้ระยะเวลาในการทำข้อสอบที่แตกต่างกัน สอดคล้องกับ Ferguson และคณะ (2002) ที่พบว่า ระยะเวลาที่มากขึ้นจะทำให้เกิดการเปลี่ยนคำตอบ ซึ่งระยะเวลาเฉลี่ยที่ใช้ในการตอบข้อสอบ 1 ข้อ คิดเป็น 60 วินาที โดยมีช่วงระยะเวลาที่ใช้ตั้งแต่ 11 วินาที ถึง 266 วินาที ดังนั้นจึงสะท้อนให้เห็นว่า การเปลี่ยนคำตอบจะเกิดขึ้นได้นั้นควรมีระยะเวลาที่เหมาะสม ถ้าน้อยเกินไปอาจไม่ทำให้เกิดการเปลี่ยนคำตอบ แต่จากการศึกษาวิจัยในเรื่องการเปลี่ยนคำตอบนั้นยังไม่ได้มีการระบุไว้ หากให้

ระยะเวลาสอบที่เพิ่มมากขึ้นจะส่งผลต่อการเปลี่ยนคำตอบหรือไม่อย่างไร ดังนั้นผู้วิจัยจึงมีความสนใจว่า ระยะเวลาในการสอบที่เหมาะสมหรือเพิ่มมากขึ้นจะทำให้เกิดการเปลี่ยนคำตอบหรือไม่อย่างไร

ทั้งนี้จากปัญหาได้ที่กล่าวมาข้างต้น ผู้วิจัยจึงมีความสนใจศึกษาผลที่ได้จากการทวนคำตอบในบริบทที่มีการเรียงลำดับของข้อสอบ ระยะเวลาในการสอบ และความสามารถของผู้สอบแตกต่างกัน บริบทใดที่จะทำให้การเปลี่ยนคำตอบเป็นประโยชน์มากที่สุด ซึ่งในการวัดผลที่ได้จากการทวนคำตอบนั้น ผู้วิจัยได้สนใจศึกษาจากค่าเฉลี่ยจำนวนข้อที่มีการทวนคำตอบที่แตกต่างกัน 5 รูปแบบ ได้แก่ การคงคำตอบไว้แล้วผิด การคงคำตอบไว้แล้วถูก การเปลี่ยนคำตอบจากถูกเป็นผิด การเปลี่ยนคำตอบจากผิดเป็นถูก การเปลี่ยนคำตอบจากผิดเป็นผิด เพื่อเป็นแนวทางสำหรับครูหรือผู้ออกข้อสอบที่จะได้นำผลวิจัยไปใช้ในการออกแบบและสร้างข้อสอบเพื่อให้สะท้อนถึงความสามารถที่แท้จริงของนักเรียนหรือผู้สอบ ตลอดจนได้ข้อมูลที่เป็นแนวทางในการบริหารจัดการสอบและให้คำแนะนำที่เป็นประโยชน์ในการทวนคำตอบแก่นักเรียนได้อย่างเหมาะสมกับบริบท

วัตถุประสงค์

1. เพื่อศึกษาความสัมพันธ์ระหว่างผลที่ได้จากการทวนคำตอบกับระดับความยากของข้อสอบ และระหว่างผลที่ได้จากการทวนคำตอบกับระดับความสามารถของผู้สอบ

2. เพื่อเปรียบเทียบผลที่ได้จากการทวนคำตอบที่มีการเรียงลำดับข้อสอบ ระยะเวลาในการสอบ และระดับความสามารถของผู้สอบที่แตกต่างกัน โดยมีวัตถุประสงค์ย่อย ดังนี้

2.1 เพื่อศึกษาปฏิสัมพันธ์ (interaction) ของปัจจัยการเรียงลำดับข้อสอบ ระยะเวลาในการสอบ และระดับความสามารถของผู้สอบที่มีต่อผลที่ได้จากการทวนคำตอบ

2.2 เพื่อศึกษาอิทธิพลหลัก (main effect) ของปัจจัยการเรียงลำดับข้อสอบ ระยะเวลาในการสอบ และระดับความสามารถของผู้สอบที่มีต่อผลที่ได้จากการทวนคำตอบ

ขอบเขตของการวิจัย

การวิจัยในครั้งนี้เป็นการวิจัยที่มุ่งเปรียบเทียบผลที่ได้จากการทวนคำตอบ ที่มีการเรียงลำดับข้อสอบ ระยะเวลาในการสอบ และระดับความสามารถของผู้สอบที่แตกต่างกัน ซึ่งการศึกษางานวิจัยที่เกี่ยวข้อง ผู้วิจัยพบว่าปัจจัยที่เกี่ยวข้องกับการทวนคำตอบซึ่งส่งผลให้มีการเปลี่ยนคำตอบ สามารถจำแนกออกเป็น 3 ด้าน คือ 1) ด้านคุณลักษณะข้อสอบ ได้แก่ การเรียงลำดับข้อสอบ 2) ด้านคุณลักษณะของผู้สอบ ได้แก่ ระดับความสามารถของผู้สอบ 3) ด้านการจัดการสอบ ได้แก่ ระยะเวลาในการสอบ ดังนั้นตัวแปรที่ศึกษาในครั้งนี้จึงประกอบด้วย

1. ตัวแปรอิสระ (independent variable) ได้แก่

1.1 การเรียงลำดับข้อสอบ

- 1) เรียงตามระดับความยากจากง่ายไปยาก
- 2) เรียงตามระดับพฤติกรรม จากขั้นต่ำไปขั้นสูง
- 3) เรียงลำดับอย่างสุ่ม

1.2 ระยะเวลาในการสอบ

- 1) 60 วินาที ต่อข้อสอบ 1 ข้อ
- 2) 75 วินาที ต่อข้อสอบ 1 ข้อ

1.3 ระดับความสามารถของผู้สอบ

- 1) ระดับความสามารถสูง
- 2) ระดับความสามารถปานกลาง
- 3) ระดับความสามารถต่ำ

2. ตัวแปรตาม (dependent variable) ได้แก่

- 2.1 ค่าเฉลี่ยจำนวนข้อที่ทวนคำตอบแล้วคงคำตอบถูกไว้
- 2.2 ค่าเฉลี่ยจำนวนข้อที่ทวนคำตอบแล้วคงคำตอบผิดไว้
- 2.3 ค่าเฉลี่ยจำนวนข้อที่ทวนคำตอบแล้วมีการเปลี่ยนคำตอบจากถูกเป็นผิด
- 2.4 ค่าเฉลี่ยจำนวนข้อที่ทวนคำตอบแล้วมีการเปลี่ยนคำตอบจากผิดเป็นถูก
- 2.5 ค่าเฉลี่ยจำนวนข้อที่ทวนคำตอบแล้วมีการเปลี่ยนคำตอบจากผิดเป็นผิด

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้ใช้ระเบียบวิธีวิจัยเชิงทดลอง ซึ่งมีวัตถุประสงค์ของการวิจัย คือ 1) เพื่อศึกษาความสัมพันธ์ระหว่างผลที่ได้จากการทวนคำตอบกับระดับความยากของข้อสอบและระหว่างผลที่ได้จากการทวนคำตอบกับระดับความสามารถของผู้สอบ 2) เพื่อเปรียบเทียบผลที่ได้จากการทวนคำตอบที่มีการเรียงลำดับข้อสอบ ระยะเวลาในการสอบและระดับความสามารถของผู้สอบที่แตกต่างกันโดยมีรายละเอียดดังต่อไปนี้

การกำหนดประชากรและตัวอย่าง

ประชากรที่ใช้ในการวิจัยครั้งนี้ คือ นักเรียนระดับชั้นมัธยมศึกษาปีที่ 4 ปีการศึกษา 2556 โรงเรียนมัธยมศึกษา ในเขตกรุงเทพมหานคร สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน (สพฐ.) จำนวน 119 โรงเรียน ซึ่งมีจำนวนนักเรียนทั้งสิ้น 42,496 คน (สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน กระทรวงศึกษาธิการ, 2556)

กลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้ ได้มาจากการสุ่มแบบหลายขั้นตอน (multistage random) โดยมีขั้นตอนการสุ่มดังนี้

ขั้นตอนที่ 1 สุ่มโดยใช้สำนักงานเขตพื้นที่การศึกษาซึ่งมีทั้งหมด 2 เขต คือ เขต 1 และเขต 2 เป็นหน่วยในการสุ่ม โดยการสุ่มอย่างง่าย (simple random sampling) ได้สำนักงานเขตพื้นที่การศึกษา มัธยมศึกษาเขต 2 ซึ่งมีโรงเรียนที่มีชั้นมัธยมศึกษาปีที่ 4 จำนวน 52 โรงเรียน รวม 21,911 คน

ขั้นตอนที่ 2 กำหนดขนาดของกลุ่มตัวอย่าง ผู้วิจัยกำหนดขนาดกลุ่มตัวอย่างนักเรียนที่กำลังศึกษาอยู่ในระดับชั้นมัธยมศึกษาปีที่ 4 ปีการศึกษา 2556 โรงเรียนมัธยมศึกษา ในเขตกรุงเทพมหานคร สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน โดยใช้โปรแกรม G*Power 3.1 ในการคำนวณเพื่อกำหนดจำนวนกลุ่มตัวอย่าง ที่มีค่าขนาดอิทธิพล (Effect size) ขนาดกลาง 0.25 (Cohen, 1977) ได้กลุ่มตัวอย่างจำนวน 197 คน แต่เพื่อป้องกันการสูญหายของข้อมูล และการวิจัยในครั้งนี้ยังอยู่ภายใต้การวิเคราะห์หาค่าพารามิเตอร์ b โดยใช้โมเดลการตอบสนองข้อสอบแบบ 1 พารามิเตอร์ ผู้วิจัยจึงเพิ่มกลุ่มตัวอย่างให้ได้จำนวนมากกว่า 197 คน เป็นจำนวน 338 คน

ขั้นตอนที่ 3 เลือกโรงเรียนในเขตพื้นที่การศึกษากรุงเทพมหานคร เขต 2 ได้โรงเรียนจำนวน 5 โรงเรียน คือ โรงเรียนนนทรีวิทยา โรงเรียนสตรีศรีสุริโยทัย โรงเรียนจันทร์หุ่นบำเพ็ญ โรงเรียนลาดปลาเค้าพิทยาคม และโรงเรียนปทุมคงคา จากนั้นเลือกห้องเรียนจากโรงเรียนทั้ง 5 โรงเรียนโดยใช้เกณฑ์ คือ ห้องเรียนที่เป็นแผนการเรียนวิทยาศาสตร์-คณิตศาสตร์ในระดับชั้นมัธยมศึกษาปีที่ 4

ขั้นตอนที่ 4 เก็บรวบรวมผลคะแนนการทดสอบ O-NET ในรายวิชา วิทยาศาสตร์ ปีการศึกษา 2555 ของนักเรียนกลุ่มตัวอย่าง เพื่อนำมาวิเคราะห์ความสามารถทางวิทยาศาสตร์ โดยจำแนกออกเป็นผู้สอบที่มีระดับความสามารถสูงจำนวน 126 คน ปานกลางจำนวน 175 คนและต่ำจำนวน 37 คน

ขั้นตอนที่ 5 สุ่มนักเรียนเข้ากลุ่มในการวิจัย โดยสุ่มอย่างง่าย (simple random sampling) ซึ่งใช้ตัวแปรอิสระเป็นเกณฑ์ในการจำแนกกลุ่มตัวอย่างเข้ากลุ่ม ได้แก่ การเรียงลำดับข้อสอบ ระยะเวลาในการสอบ และระดับความสามารถ

เครื่องมือที่ใช้ในการวิจัย

1. แบบสอบแบบหลายตัวเลือก รายวิชา ชีววิทยา เรื่อง ระบบย่อยอาหาร จำนวน 70 ข้อ ประกอบด้วยจำนวนตัวเลือก 4 ตัวเลือก สำหรับคัดเลือกเพื่อวิเคราะห์หาค่าพารามิเตอร์ของข้อสอบตามทฤษฎีการตอบสนองข้อสอบแบบ 1 พารามิเตอร์ ซึ่งมีค่าความเที่ยงเป็น 0.87

2. แบบสอบแบบหลายตัวเลือก รายวิชา ชีววิทยา เรื่อง ระบบย่อยอาหาร จำนวน 35 ข้อ ที่ผ่านการคัดเลือกแล้ว โดยมีค่าความยาก (b) อยู่ระหว่าง -2.50 ถึง +2.50 จำนวน 3 ชุด ได้แก่ ชุดที่ 1 เรียงลำดับข้อสอบตามระดับความยากจากระดับง่ายไปยาก ชุดที่ 2 เรียงลำดับข้อสอบตามระดับพฤติกรรมจากขั้นต่ำไปขั้นสูง และชุดที่ 3 เรียงลำดับข้อสอบอย่างสุ่ม

การเก็บรวบรวมข้อมูล

การวิจัยในครั้งนี้มีขั้นตอนการเก็บข้อมูลแบ่งออกเป็น 4 ขั้นตอน โดยมีขั้นตอนการดำเนินการ ดังนี้

ขั้นตอนที่ 1 การนำข้อสอบไปทดลองใช้เพื่อวิเคราะห์หาค่า b กับนักเรียนจำนวน 222 คน ซึ่งเป็นนักเรียนในระดับชั้นมัธยมศึกษาปีที่ 4 โรงเรียนศรีพฤฒา โรงเรียนสาธิตจุฬาลงกรณ์มหาวิทยาลัย และโรงเรียนสตรีศรีสุริโยทัย ผลการวิเคราะห์คุณภาพของข้อสอบพบว่ามีค่าความเที่ยงเท่ากับ 0.8789 ซึ่งถือว่าอยู่ในเกณฑ์ดี นอกจากนี้ผู้วิจัยยังได้นำผลการวิเคราะห์ค่าความยาก (b) ที่ได้จากการวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบแบบ 1 พารามิเตอร์โดยใช้โปรแกรม MULTILOG 7.03 มาใช้เพื่อคัดเลือกข้อสอบให้เหลือ 35 ข้อ แล้วนำมาจัดเรียงข้อสอบตามค่าความยาก

ขั้นตอนที่ 2 การคัดเลือกข้อสอบและนำมาจัดเรียงได้ข้อสอบจำนวน 35 ข้อ จำนวน 3 ชุด ได้แก่ ชุดที่ 1 เรียงลำดับข้อสอบตามระดับความยากจากระดับง่ายไปยาก ชุดที่ 2 เรียงลำดับข้อสอบตามระดับพฤติกรรมจากขั้นต่ำไปขั้นสูง ชุดที่ 3 เรียงลำดับข้อสอบอย่างสุ่ม

ขั้นตอนที่ 3 การวิเคราะห์ผลคะแนนการทดสอบ O-NET วิชา วิทยาศาสตร์เพื่อจัดจำแนกผู้สอบตามระดับความสามารถทางวิทยาศาสตร์

ขั้นตอนที่ 4 การวิเคราะห์หาความสัมพันธ์ระหว่างระดับความยากของข้อสอบและระดับความสามารถของผู้สอบกับผลที่ได้จากการทวนคำตอบ และการวิเคราะห์เพื่อเปรียบเทียบผลที่ได้จากการทวนคำตอบ ที่มีการเรียงลำดับข้อสอบ ระยะเวลาในการสอบ และระดับความสามารถของผู้สอบแตกต่างกัน

การวิเคราะห์ข้อมูล

การวิเคราะห์ข้อมูลเพื่อศึกษาความสัมพันธ์ระหว่างระดับความยากของข้อสอบกับผลที่ได้จากการทวนคำตอบ ความสัมพันธ์ระหว่างระดับความสามารถของผู้สอบกับผลที่ได้จากการทวนคำตอบและเปรียบเทียบผลที่ได้จากการทวนคำตอบ โดยผู้วิจัยได้แบ่งออกเป็น 4 ขั้นตอน คือ

1. การวิเคราะห์ค่าพารามิเตอร์โดยใช้โปรแกรม MULTILOG 7.03 เพื่อคัดเลือกข้อสอบที่ดี โดยมีค่าความยาก (b) อยู่ระหว่าง -2.50 ถึง +2.50 แล้วคัดเลือกให้เหลือเพียง 35 ข้อ จากทั้งหมด 70 ข้อ

2. การวิเคราะห์ผลคะแนนการทดสอบ O-NET วิชา วิทยาศาสตร์ระดับชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2555 ระดับจังหวัดเพื่อจัดจำแนกนักเรียนตามระดับความสามารถทางวิทยาศาสตร์ ออกเป็น 3 กลุ่ม คือ กลุ่มสูง ปานกลาง ต่ำ โดยใช้วิธีการแบ่งกลุ่มแบบอิงกลุ่ม (norm-referenced evaluation)

ภายใต้การใช้ค่าเฉลี่ยกับส่วนเบี่ยงเบนมาตรฐาน ในการวิเคราะห์เพื่อจัดจำแนกความสามารถทางวิทยาศาสตร์ของนักเรียน

3. การวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรต่างๆกับผลที่ได้จากการทวนคำตอบ โดยการวิเคราะห์ความสัมพันธ์ระหว่างผลที่ได้จากการทวนคำตอบกับระดับความยากของข้อสอบ ใช้การวิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน (Pearson Product Moment Correlation Coefficient) และการวิเคราะห์ความสัมพันธ์ระหว่างผลที่ได้จากการทวนคำตอบกับระดับความสามารถของผู้สอบ ใช้การวิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมน (Spearman's rank Correlation)

4.การวิเคราะห์เพื่อเปรียบเทียบผลที่ได้จากการทวนคำตอบที่มีการเรียงลำดับข้อสอบ ระยะเวลาในการสอบและระดับความสามารถของผู้สอบที่แตกต่างกัน ใช้การวิเคราะห์ความแปรปรวนสามทาง (Three-Way ANOVA) โดยใช้โปรแกรมคอมพิวเตอร์เพื่อเปรียบเทียบความแตกต่างของผลที่ได้จากการทวนคำตอบ 5 รูปแบบ ภายใต้ตัวแปรต้น คือ การเรียงลำดับข้อสอบ (เรียงตามระดับความยาก,เรียงตามระดับพฤติกรรมและเรียงอย่างสุ่ม)ระยะเวลาในการสอบ (60 วินาทีต่อข้อ และ 75 วินาทีต่อข้อ) และระดับความสามารถของผู้สอบ (สูง , ปานกลาง และต่ำ) ซึ่งแบ่งการวิเคราะห์ออกเป็น 5 ครั้งตามจำนวนของตัวแปรตาม คือ ผลที่ได้จากการทวนคำตอบ ซึ่งมี 5 รูปแบบย่อย ดังนี้ 1) ค่าเฉลี่ยจำนวนข้อที่ทวนคำตอบแล้วยังคงคำตอบถูกไว้ (RR) 2) ค่าเฉลี่ยจำนวนข้อที่ทวนคำตอบแล้วยังคงคำตอบผิด (WW) 3) ค่าเฉลี่ยจำนวนข้อที่ทวนคำตอบแล้วเปลี่ยนคำตอบจากถูกเป็นผิด (R*W) 4) ค่าเฉลี่ยจำนวนข้อที่ทวนคำตอบแล้วเปลี่ยนคำตอบจากผิดเป็นถูก (W*R) 5) ค่าเฉลี่ยจำนวนข้อที่ทวนคำตอบแล้วเปลี่ยนคำตอบจากผิดเป็นผิด (W*W)

ผลการวิจัย

การวิจัยเรื่องการเปรียบเทียบผลที่ได้จากการทวนคำตอบในแบบสอบแบบหลายตัวเลือกที่มีการเรียงลำดับข้อสอบ ระยะเวลาในการสอบ และระดับความสามารถของผู้สอบแตกต่างกันสรุปผลการวิจัย ดังนี้

1.ผลการวิเคราะห์ความสัมพันธ์ระหว่างผลที่ได้จากการทวนคำตอบกับระดับความยากของข้อสอบ

จากการวิเคราะห์ความสัมพันธ์ระหว่างผลที่ได้จากการทวนคำตอบทั้ง 5 รูปแบบพบว่า ผลที่ได้จากการทวนคำตอบมีความสัมพันธ์กับระดับความยากของข้อสอบจำนวน 3 คู่ ได้แก่ 1) การคงคำตอบไว้แล้วถูก (RR) มีความสัมพันธ์กับระดับความยากของข้อสอบ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ($R = -.696$) 2) การคงคำตอบไว้แล้วผิด (WW) มีความสัมพันธ์กับระดับความยากของข้อสอบอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ($R = .683$) 3) การเปลี่ยนคำตอบจากผิดเป็นถูก (W*R) มีความสัมพันธ์กับระดับความยากของข้อสอบอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ($R = -.316$) ดังตาราง 1

ตาราง 1 ผลการวิเคราะห์ความสัมพันธ์ระหว่างผลที่ได้จากการทวนคำตอบกับระดับความยากของข้อสอบ

ตัวแปร	ผลที่ได้จากการทวนคำตอบ				
	RR	WW	R*W	W*R	W*W
ระดับความยากของข้อสอบ (b)	-0.696**	0.683**	0.154	-0.316*	-0.080

2) ผลการวิเคราะห์ความสัมพันธ์ระหว่างผลที่ได้จากการทวนคำตอบกับระดับความสามารถของผู้สอบ

จากการวิเคราะห์ความสัมพันธ์ระหว่างผลที่ได้จากการทวนคำตอบทั้ง 5 รูปแบบพบว่า ผลที่ได้จากการทวนคำตอบมีความสัมพันธ์กับระดับความสามารถของผู้สอบ จำนวน 2 คู่ ได้แก่ 1) การคงคำตอบไว้แล้วถูก (RR) มีความสัมพันธ์กับความสามารถของผู้สอบอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ($R = .424$)

2) การคงคำตอบไว้แล้วผิด (WW) มีความสัมพันธ์กับระดับความสามารถของผู้สอบอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ($R = -.407$) ดังตาราง 2

ตาราง 2 ผลการวิเคราะห์ความสัมพันธ์ระหว่างผลที่ได้จากการทวนคำตอบกับระดับความสามารถของผู้สอบ

ตัวแปร	ผลที่ได้จากการทวนคำตอบ				
	RR	WW	R*W	W*R	W*W
ระดับความสามารถ (Ability)	.424**	-.407**	.066	.031	-.024

3) ปัจจัยการเรียงลำดับข้อสอบ ระยะเวลาในการสอบและระดับความสามารถของผู้สอบ (ORDER * RESPONSE TIME * ABILITY) ไม่มีผลร่วมกันต่อค่าเฉลี่ยจำนวนข้อที่ทวนคำตอบแล้วยังคงคำตอบถูกไว้ (RR) อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ($F = 1.924, sig = 0.106$) นอกจากนี้ผลการศึกษาก็พบหลัก (main effect) พบว่า 1) ระยะเวลาในการสอบ (RESPONSE TIME) มีผลต่อค่าเฉลี่ยจำนวนข้อที่ทวนคำตอบแล้วยังคงคำตอบถูกไว้ (RR) อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 2) ระดับความสามารถของผู้สอบ (ABILITY) มีผลต่อค่าเฉลี่ยจำนวนข้อที่ทวนคำตอบแล้วยังคงคำตอบถูกไว้ (RR) อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 3) การเรียงลำดับข้อสอบ (ORDER) ไม่มีผลต่อค่าเฉลี่ยจำนวนข้อที่ทวนคำตอบแล้วยังคงคำตอบถูกไว้ (RR) อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

4) ปัจจัยการเรียงลำดับข้อสอบ ระยะเวลาในการสอบและระดับความสามารถของผู้สอบ (ORDER * RESPONSE TIME * ABILITY) ไม่มีผลร่วมกันต่อค่าเฉลี่ยจำนวนข้อที่ทวนคำตอบแล้วยังคงคำตอบผิดไว้ (WW) อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ($F = 1.023, sig = 0.395$) นอกจากนี้ผลการศึกษาก็พบหลัก (main effect) พบว่า 1) ระยะเวลาในการสอบ (RESPONSE TIME) มีผลต่อค่าเฉลี่ยจำนวนข้อที่ทวนคำตอบแล้วยังคงคำตอบผิดไว้ (WW) อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ($F = 33.80, sig = .000$) 2) ระดับความสามารถของผู้สอบมีผลต่อค่าเฉลี่ยจำนวนข้อที่ทวนคำตอบแล้วยังคงคำตอบผิดไว้ (WW) อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ($F = 27.22, sig = .000$) 3) การเรียงลำดับข้อสอบ (ORDER) ไม่มีผลต่อค่าเฉลี่ยจำนวนข้อที่ทวนคำตอบแล้วยังคงคำตอบผิดไว้ (WW) อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ($F = 0.290, sig = .744$)

5) การเรียงลำดับข้อสอบ ระยะเวลาในการสอบและระดับความสามารถของผู้สอบ (ORDER * RESPONSE TIME * ABILITY) ไม่มีผลร่วมกันต่อค่าเฉลี่ยของจำนวนข้อที่ทวนคำตอบแล้วเปลี่ยนจากถูกเป็นผิด (R*W) ในแบบสอบแบบหลายตัวเลือก ($F = 2.048, sig = 0.088$) แต่ระยะเวลาในการสอบ และระดับความสามารถของผู้สอบ (RESPONSE TIME * ABILITY) มีผลร่วมกันต่อค่าเฉลี่ยของจำนวนข้อที่ทวนคำตอบแล้วเปลี่ยนจากถูกเป็นผิด (R*W) ในแบบสอบแบบหลายตัวเลือกอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ($F = 5.835, sig = 0.003$) โดยปัจจัยระยะเวลาในการสอบ 75 วินาทีต่อข้อ พบว่าในผู้สอบที่มีระดับความสามารถแตกต่างกัน มีจำนวนข้อที่ทวนคำตอบแล้วเปลี่ยนคำตอบจากถูกเป็นผิดแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ($F = 5.424, sig = .005$) ดังตาราง 3

ตาราง 3 ผลการวิเคราะห์ ความต่างของค่าเฉลี่ยของจำนวนข้อที่ทวนคำตอบแล้วเปลี่ยนจากถูกเป็นผิด (R*W) ในแบบสอบแบบหลายตัวเลือกเมื่อระยะเวลาในการสอบและระดับความสามารถของผู้สอบแตกต่างกัน

ระยะเวลาในการสอบ (วินาที/ข้อ)	ระดับความสามารถของผู้สอบ	จำนวนผู้สอบ (คน)	จำนวนข้อสอบ (ข้อ)				Levene's test		F-test	
			Min	Max	Mean	SD	F	Sig.	F	Sig.
60	ต่ำ	14	0	1	0.14	0.36	5.598	0.004	1.360	0.259
	ปานกลาง	72	0	4	0.29	0.77				
	สูง	78	0	1	0.14	0.35				
	รวม	164	0	4	0.21	0.58				
75	ต่ำ	23	0	3	0.30	0.70	13.780	0.000	5.424*	0.005
	ปานกลาง	103	0	1	0.16	0.36				
	สูง	48	0	3	0.44	6.16				
	รวม	174	0	3	0.25	0.51				

*p < .05

6) ระยะเวลาในการสอบ และระดับความสามารถของผู้สอบ (RESPONSE TIME * ABILITY) มีผลร่วมกันต่อค่าเฉลี่ยของจำนวนข้อที่ทวนคำตอบแล้วเปลี่ยนจากผิดเป็นถูก (W*R) ในแบบสอบแบบหลายตัวเลือกอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 โดยปัจจัยระยะเวลาในการสอบ 75 วินาทีต่อข้อ พบว่าในผู้สอบที่มีระดับความสามารถแตกต่างกัน มีจำนวนข้อที่ทวนคำตอบแล้วเปลี่ยนคำตอบจากผิดเป็นถูกแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ($F = 4924$, $sig = .008$) รายละเอียดดังตาราง 4

ตาราง 4 ผลการวิเคราะห์ค่าเฉลี่ยของจำนวนข้อที่ทวนคำตอบแล้วเปลี่ยนจากผิดเป็นถูก (W*R) ในแบบสอบแบบหลายตัวเลือกเมื่อระยะเวลาในการสอบและระดับความสามารถของผู้สอบแตกต่างกัน

ระยะเวลาในการสอบ (วินาที/ข้อ)	ระดับความสามารถของผู้สอบ	จำนวนผู้สอบ (คน)	จำนวนข้อสอบ (ข้อ)				Levene's test		F-test	
			Min	Max	Mean	SD	F	Sig.	F	Sig.
60	ต่ำ	14	0	2	0.43	0.64	3.713	0.027	1.546	0.216
	ปานกลาง	72	0	4	0.57	0.85				
	สูง	78	0	2	0.36	0.62				
	รวม	164	0	4	0.46	0.73				
75	ต่ำ	23	0	2	0.39	0.66	10.620	0.000	4.924*	0.008
	ปานกลาง	103	0	3	0.25	0.55				
	สูง	48	0	5	0.67	1.10				
	รวม	174	0	5	0.39	0.77				

*p < .05

7) ปัจจัยการเรียงลำดับข้อสอบ ระยะเวลาในการสอบและระดับความสามารถของผู้สอบที่ (ORDER * RESPONSE TIME * ABILITY) ไม่มีผลร่วมกันต่อค่าเฉลี่ยจำนวนจำนวนข้อที่ทวนคำตอบแล้วมีการเปลี่ยนคำตอบจากผิดเป็นผิด (W*W) อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ($F = 0.362$, $sig = 0.836$) นอกจากนี้ยังพบว่า การเรียงลำดับข้อสอบ (ORDER) ไม่มีผลต่อค่าเฉลี่ยจำนวนข้อที่ทวนคำตอบแล้วมีการเปลี่ยนคำตอบจากผิดเป็นผิด (W*W) อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ($F = 0.035$, $sig = .966$) ระยะเวลาในการสอบ (RESPONSE TIME) ไม่มีผลต่อค่าเฉลี่ยจำนวนข้อที่ทวนคำตอบแล้วมีการเปลี่ยนคำตอบจากผิดเป็นผิด (W*W)

อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ($F = 0.364$, $sig = .547$) และระดับความสามารถของผู้สอบ (ABILITY) ไม่ส่งผลต่อค่าเฉลี่ยจำนวนข้อที่ทวนคำตอบแล้วมีการเปลี่ยนคำตอบจากผิดเป็นผิด (W^*W) อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ($F = 0.566$, $sig = .568$)

อภิปรายผลและข้อเสนอแนะ

1. การวิเคราะห์ความสัมพันธ์ระหว่างผลที่ได้จากการทวนคำตอบกับระดับความยากของข้อสอบ

ผลการวิเคราะห์โดยภาพรวมพบว่า การทวนคำตอบแล้วยังคงคำตอบถูกไว้ (RR) มีความสัมพันธ์กับระดับความยากของข้อสอบ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ($R = -.696$) แสดงให้เห็นว่า ข้อสอบที่มีระดับความยากสูงขึ้นจะยังมีการคงคำตอบถูกไว้ (RR) ลดลง การทวนคำตอบแล้วยังคงคำตอบผิดไว้ (W) มีความสัมพันธ์กับระดับความยากของข้อสอบอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ($R = .683$) แสดงให้เห็นว่า ข้อสอบที่มีระดับความยากสูงขึ้นจะยังมีการคงคำตอบผิดไว้ (W) สูงขึ้น และการเปลี่ยนคำตอบจากผิดเป็นถูก (W^*R) มีความสัมพันธ์กับระดับความยากของข้อสอบอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ($R = -.316$) แสดงให้เห็นว่า ข้อสอบที่มีระดับความยากสูงขึ้นจะยังมีการเปลี่ยนคำตอบจากผิดเป็นถูกลดลง สอดคล้องกับงานวิจัยของ Al-Hamly & Coombe (2005) ที่พบว่าระดับความยากของข้อสอบมีความสัมพันธ์ในทิศทางบวก ($r = .217$) กับจำนวนของการเปลี่ยนคำตอบจากผิดเป็นถูกอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และ Ferguson, Kreiter, Peterson, Rowat, & Elliott (2002) ที่พบว่า ค่าความยาก มีความสัมพันธ์กับการเปลี่ยนคำตอบระหว่างการสอบอย่างมีนัยสำคัญทางสถิติ ($r = -.70$, $p < .0001$)

2. การวิเคราะห์ความสัมพันธ์ระหว่างผลที่ได้จากการทวนคำตอบกับระดับความสามารถของผู้สอบ

ผลการวิเคราะห์โดยภาพรวมพบว่า การทวนคำตอบแล้วยังคงคำตอบถูกไว้ (RR) มีความสัมพันธ์กับความสามารถของผู้สอบ (Ability) อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ($R = .424$) แสดงให้เห็นว่า นักเรียนที่มีระดับความสามารถสูงจะคงคำตอบถูกไว้ (R) มากกว่านักเรียนกลุ่มปานกลางและกลุ่มต่ำ และการทวนคำตอบแล้วยังคงคำตอบผิดไว้ (W) มีความสัมพันธ์กับระดับความสามารถของผู้สอบ (Ability) อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ($R = -.407$) แสดงให้เห็นว่า นักเรียนที่มีระดับความสามารถสูงจะคงคำตอบผิดไว้น้อยกว่านักเรียนกลุ่มปานกลางและกลุ่มต่ำ สอดคล้อง ขวัญใจ อุตราภรณ์ (2528) ที่พบว่าผู้สอบที่มีความสามารถแตกต่างกันจะมีรูปแบบการเปลี่ยนคำตอบต่างกัน นอกจากนี้ยังสอดคล้องกับงานวิจัยของ พิมพ์สิริ เจริญนเรศเศรษฐ์ (2549) ที่พบว่า ผู้สอบที่ทวนคำตอบแล้วยังคงคำตอบเดิมไว้มากที่สุดคือ ผู้สอบที่มีระดับความสามารถสูง (48.61%) แต่ไม่ได้มีการจำแนกว่าการทวนคำตอบแล้วคงคำตอบถูกและคงคำตอบผิดไว้อย่างไรในผู้สอบแต่ละระดับความสามารถ

3. การเปรียบเทียบผลที่ได้จากการทวนคำตอบที่มีการเรียงลำดับข้อสอบ ระยะเวลาในการสอบ และระดับความสามารถของผู้สอบที่แตกต่างกัน

3.1 ค่าเฉลี่ยจำนวนข้อที่ทวนคำตอบแล้วยังคงคำตอบถูกไว้ (RR) ผลการศึกษาพบว่า **ระยะเวลาในการสอบ** ส่งผลต่อค่าเฉลี่ยจำนวนข้อที่ทวนคำตอบแล้วยังคงคำตอบถูกไว้ (RR) อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 โดยระยะเวลาในการสอบ 60 วินาทีต่อข้อ มีค่าเฉลี่ยจำนวนข้อที่ทวนคำตอบแล้วยังคงคำตอบถูกไว้ (R) มากกว่าระยะเวลาในการสอบ 75 วินาทีต่อข้อ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ($F = 43.04$, $sig = .000$) สอดคล้องกับงานวิจัยของ Ferguson และคณะ (2002) ที่พบว่า ระยะเวลาที่ใช้ในการตอบข้อสอบนั้นมีความสัมพันธ์กับการเปลี่ยนคำตอบซึ่งระบุว่าระยะเวลาที่มากขึ้นจะทำให้เกิดการเปลี่ยนคำตอบ นอกจากนี้ **ด้านระดับความสามารถของผู้สอบ**พบว่า นักเรียนที่มีระดับความสามารถสูงมีค่าเฉลี่ยจำนวนข้อที่ทวนคำตอบแล้วยังคงคำตอบถูกไว้ (RR) มากกว่านักเรียนที่มีระดับความสามารถปานกลางและ

นักเรียนกลุ่มต่ำ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 สอดคล้องกับงานวิจัยของ พิมพ์สิริ เจริญนเรศวร (2549) ที่พบว่า ผู้สอบที่ทวนคำตอบแล้วยังคงคำตอบเดิมไว้มากที่สุดคือ ผู้สอบที่มีระดับความสามารถสูง (48.61%) นอกจากนี้ยังพบว่า**ด้านการเรียงลำดับข้อสอบ** ไม่ส่งผลต่อค่าเฉลี่ยจำนวนข้อที่ทวนคำตอบแล้วยังคงคำตอบถูกไว้ (RR) อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แสดงให้เห็นว่าการเรียงลำดับข้อสอบไม่ส่งผลต่อการทวนคำตอบ เนื่องจากในการวิจัยครั้งนี้ผู้วิจัยได้ออกแบบการเรียงลำดับข้อสอบอย่างเป็นระบบ คือ เรียงลำดับจากง่ายไปยาก เรียงลำดับตามระดับพฤติกรรมขั้นต่ำไปขั้นสูง และเรียงลำดับอย่างสุ่มภายใต้เนื้อหาเดียวกัน ตามผลการวิจัยที่ ศิริวิมล สุขสมบูรณ์ (2541) ได้แนะนำไว้ว่า รูปแบบการเรียงลำดับข้อสอบแบบเรียงจากข้อง่ายไปหาข้อยากตามเนื้อหาของบทเรียนจึงเป็นรูปแบบการจัดเรียงข้อคำถามที่เหมาะสมที่สุด

3.2 ค่าเฉลี่ยจำนวนข้อที่ทวนคำตอบแล้วยังคงคำตอบผิดไว้ (WW) ผลการศึกษาพบว่า มีปฏิสัมพันธ์กันระหว่าง การเรียงลำดับข้อสอบ ระยะเวลาในการสอบ และระดับความสามารถของผู้สอบ ($ORDER * RESPONSE TIME * ABILITY$) อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ($F = 1.023, sig = 0.395$) หรือมีค่าเฉลี่ยจำนวนข้อที่ทวนคำตอบแล้วยังคงคำตอบผิดไว้ (WW) ไม่แตกต่างกันเมื่อมีการเรียงลำดับข้อสอบ ระยะเวลาในการสอบ และระดับความสามารถของผู้สอบแตกต่างกัน นอกจากนี้ **ด้านระยะเวลาในการสอบ** พบว่า ระยะเวลา 75 วินาทีต่อข้อ มีค่าเฉลี่ยจำนวนข้อที่ทวนคำตอบแล้วยังคงคำตอบผิดไว้ (WW) สูงกว่า 60 วินาทีต่อข้อ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ($F = 33.80, sig = .000$) แสดงให้เห็นว่าเมื่อมีการเพิ่มระยะเวลาในการสอบมากขึ้นจาก 60 เป็น 75 วินาที ต่อข้อ มีค่าเฉลี่ยจำนวนข้อที่ทวนคำตอบแล้วยังคงคำตอบผิด(WW)สูงขึ้น **ด้านระดับความสามารถของผู้สอบ**พบว่า นักเรียนที่มีระดับความสามารถต่ำมีค่าเฉลี่ยจำนวนข้อที่ทวนคำตอบแล้วยังคงคำตอบผิดไว้ (WW) มากกว่า นักเรียนที่มีระดับความสามารถปานกลางและนักเรียนกลุ่มสูง อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และ**ด้านการเรียงลำดับข้อสอบ** ไม่ส่งผลต่อค่าเฉลี่ยจำนวนข้อที่ทวนคำตอบแล้วยังคงคำตอบผิดไว้ (WW) อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ($F = 0.290, sig = .744$) แสดงให้เห็นว่า การเรียงลำดับข้อสอบไม่ส่งผลต่อการทวนคำตอบ

3.3 ค่าเฉลี่ยจำนวนข้อที่ทวนคำตอบแล้วมีการเปลี่ยนคำตอบจากถูกเป็นผิด ($R*W$) พบว่ามีค่าเฉลี่ยจำนวนข้อที่ทวนคำตอบแล้วมีการเปลี่ยนคำตอบจากถูกเป็นผิด ($R*W$) แตกต่างกันเมื่อมีระยะเวลาในการสอบ และระดับความสามารถของผู้สอบแตกต่างกันโดย ระยะเวลาในการสอบที่มากขึ้นคือ 75 นาทีต่อข้อ จะส่งผลให้นักเรียนที่มีระดับความสามารถสูงมีค่าเฉลี่ยจำนวนข้อที่ทวนคำตอบแล้วเปลี่ยนจากถูกเป็นผิด ($R*W$) ($M = 0.44, SD = 6.16$) ซึ่งมากกว่านักเรียนที่มีระดับความสามารถปานกลาง ($M = 0.16, SD = 0.36$) อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แสดงให้เห็นว่าเมื่อเพิ่มระยะเวลาในการสอบมากขึ้นนักเรียนกลุ่มสูงจะทวนแล้วเปลี่ยนคำตอบจากถูกเป็นผิดมากกว่านักเรียนกลุ่มปานกลางที่เป็นเช่นนั้นน่าจะมีสาเหตุจากการที่นักเรียนกลุ่มสูงจะมีความคิดรอบคอบในการทำข้อสอบและทวนคำตอบมากกว่าเด็กกลุ่มกลางและต่ำ จึงมีการทวนและเปลี่ยนคำตอบเพื่อแก้ไขข้อที่ตนเองไม่แน่ใจอีกครั้ง ทำให้มีการเปลี่ยนคำตอบมากกว่ากลุ่มกลางแต่มีรูปแบบในการเปลี่ยนคือเปลี่ยนจากถูกเป็นผิด ($R*W$) ทั้งนี้อาจมาจากระยะเวลาที่มากขึ้นจึงเกิดความลังเลในการเปลี่ยนคำตอบจากที่ถูกแล้วกลายเป็นผิด

3.4 ค่าเฉลี่ยจำนวนข้อที่ทวนคำตอบแล้วมีการเปลี่ยนคำตอบจากผิดเป็นถูก ($W*R$) พบว่า มีค่าเฉลี่ยจำนวนข้อที่ทวนคำตอบแล้วมีการเปลี่ยนคำตอบจากถูกเป็นผิด ($R*W$) แตกต่างกันเมื่อมีระยะเวลาในการสอบ และระดับความสามารถของผู้สอบแตกต่างกัน โดยปัจจัยระยะเวลาในการสอบ 75 วินาทีต่อข้อ นักเรียนที่มีระดับความสามารถสูงมีค่าเฉลี่ยจำนวนข้อที่ทวนคำตอบแล้วเปลี่ยนจากผิดเป็นถูก ($W*R$) มากกว่านักเรียนที่มีระดับความสามารถปานกลาง อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ($sig = .048$) แสดงให้เห็นว่าเมื่อเพิ่มระยะเวลาในการสอบมากขึ้นนักเรียนกลุ่มสูงจะทวนแล้วเปลี่ยนคำตอบจากผิดเป็นถูกมากกว่านักเรียน

กลุ่มปานกลางที่เป็นเช่นนี้ น่าจะมีสาเหตุจากการที่นักเรียนกลุ่มสูงจะมีความคิดรอบคอบในการทำข้อสอบและ
ทวนคำตอบมากกว่าเด็กกลุ่มกลางและต่ำ จึงมีการทวนและเปลี่ยนคำตอบเพื่อแก้ไขข้อที่ตนเองไม่แน่ใจอีกครั้ง
ทำให้มีการเปลี่ยนคำตอบมากกว่ากลุ่มกลาง ซึ่งขัดแย้งกับงานวิจัยของพิมพ์สิริ เจริญรณรงค์ (2540) ที่พบว่า
ผู้สอบที่มีระดับความสามารถปานกลาง มีการเปลี่ยนคำตอบจากผิดเป็นถูกมากที่สุด (2.24 ข้อ) และน้อยที่สุด
คือผู้สอบที่มีระดับความสามารถต่ำ (1.39 ข้อ)

3.5 ค่าเฉลี่ยจำนวนข้อที่ทวนคำตอบแล้วมีการเปลี่ยนคำตอบจากผิดเป็นผิด (W^*W) มี
ค่าเฉลี่ยจำนวนข้อที่ทวนคำตอบแล้วมีการเปลี่ยนคำตอบจากผิดเป็นผิด (W^*W) ไม่แตกต่างกันเมื่อมี
ระยะเวลาในการสอบ และระดับความสามารถของผู้สอบแตกต่างกัน โดย พบว่า **ด้านการเรียงลำดับข้อสอบ**
ด้านระยะเวลาในการสอบ และด้านระดับความสามารถของผู้สอบ ไม่ส่งผลต่อค่าเฉลี่ยจำนวนข้อที่ทวน
คำตอบแล้วมีการเปลี่ยนคำตอบจากผิดเป็นผิด (W^*W) อย่างมีนัยสำคัญทางสถิติที่ระดับ แสดงให้เห็นว่าไม่ว่า
ผู้สอบจะมีความสามารถระดับใดก็จะมี การเปลี่ยนคำตอบจากผิดเป็นไม่แตกต่างกัน ซึ่งขัดแย้งกับผลการวิจัย
ของ พิมพ์สิริ เจริญรณรงค์ (2540) ที่พบว่า ผู้สอบที่มีระดับความสามารถต่ำ มีการเปลี่ยนคำตอบจากผิดเป็น
ผิดมากที่สุด (1.88 ข้อ) และน้อยที่สุดคือผู้สอบที่มีระดับความสามารถสูง (1.11 ข้อ)

ข้อเสนอแนะในการนำไปใช้

1. ด้านการสร้างข้อสอบแบบเลือกตอบ เพื่อให้สะท้อนถึงความสามารถที่แท้จริงของนักเรียนหรือ
ผู้สอบ โดยรูปแบบที่เหมาะสมคือ การเรียงลำดับข้อสอบจากง่ายไปยาก การเรียงลำดับตามระดับพฤติกรรมชั้น
ต่ำไปชั้นสูง และการเรียงลำดับอย่างสุ่ม ตามเนื้อหาบทเรียน เนื่องจากแบบสอบแบบเลือกตอบมักมีข้อจำกัด
คือ ผู้สอบสามารถลอกคำตอบกันได้ง่ายกว่าแบบสอบแบบเขียนตอบ ดังนั้นจึงมีเทคนิคที่ครูหรือผู้สร้างข้อสอบ
ใช้ในการป้องกันการลอกข้อสอบคือ การเพิ่มจำนวนชุดของข้อสอบ โดยภายในข้อสอบแต่ละชุดจะมีลำดับ
ข้อสอบที่สลับกัน ซึ่งจะช่วยป้องกันการลอกข้อสอบหรือการโกงกันได้ แต่การนำข้อสอบมาสลับลำดับข้อ
คำถามนั้น ครูหรือผู้สร้างข้อสอบอาจไม่ได้คำนึงถึงลำดับความยากง่ายของข้อสอบ หรือลำดับของเนื้อหา ซึ่ง
ส่งผลต่อความสามารถที่แท้จริงของผู้สอบได้) โดยจากงานวิจัยที่เกี่ยวข้องกับการเรียงลำดับข้อสอบนั้นพบว่า
การเรียงลำดับข้อสอบที่เหมาะสมและสะท้อนความสามารถที่แท้จริงของผู้สอบได้ดีที่สุด คือ การเรียงลำดับ
ข้อสอบจากข้อง่ายไปยังข้อยากตามเนื้อหาของ

2. ด้านระยะเวลาที่ใช้ในการสอบ พบว่า ระยะเวลาที่เหมาะสมสำหรับผู้สอบทุกระดับความสามารถ
คือ 60 วินาทีต่อข้อ ผู้วิจัยจึงเสนอแนะว่า ในการนำผลวิจัยไปใช้ในครั้งต่อไป ควรเลือกใช้ระยะเวลาในการสอบ
ที่เหมาะสมคือ 60 วินาทีต่อข้อ เพราะหากเพิ่มระยะเวลาในการสอบมากขึ้นจะมีผลในด้านลบกับผู้สอบในทุก
ระดับความสามารถ โดยเฉพาะผู้สอบที่มีระดับความสามารถสูงที่มีการเปลี่ยนคำตอบมากขึ้นโดยเฉพาะการ
เปลี่ยนจากถูกเป็นผิด (R^*W) และจากผิดเป็นถูก (W^*R)

ข้อเสนอแนะในการวิจัยครั้งต่อไป

1. การวิจัยในครั้งนี้เป็นการเปรียบเทียบผลที่ได้จากการทวนคำตอบ ในแบบสอบแบบหลายตัวเลือก
ที่มีการเรียงลำดับข้อสอบ ระยะเวลาในการสอบและระดับความสามารถของผู้สอบที่แตกต่างกันซึ่งศึกษา
ภายใต้ข้อสอบแบบเลือกตอบหลายตัวเลือกที่เป็นแบบเขียนตอบ (Paper-Pencil Test : PPT) ที่กำหนดให้
ผู้สอบตอบโดยใช้การเขียน โดยแบบสอบแบบเขียนตอบที่ผู้วิจัยศึกษาเป็นแบบสอบแบบปรนัย แต่สำหรับใน
การวิจัยครั้งต่อไปควรเลือกใช้แบบทดสอบในระบบคอมพิวเตอร์ (Computer-Base Test : CBT) ซึ่งเป็นแบบ
สอบที่ทดสอบผ่านระบบคอมพิวเตอร์ โดยมีวัตถุประสงค์เพื่อช่วยให้สามารถประมวลผลการสอบรวดเร็วขึ้น
และลดความผิดพลาดในการระบายคำตอบลงในกระดาษคำตอบผิด ซึ่งงานวิจัยที่ผ่านมาได้ศึกษาการทวน

คำตอบโดยทดสอบผ่านระบบคอมพิวเตอร์ สามารถแบ่งเป็น 2 แบบ คือ การทดสอบทั่วไปและการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบ

2. การวิจัยในครั้งนี้เป็นการเปรียบเทียบผลที่ได้จากการทวนคำตอบ ในข้อสอบแบบหลายตัวเลือก ที่มีการเรียงลำดับข้อสอบแตกต่างกัน เนื่องจากข้อสอบแบบเลือกตอบมักมีข้อจำกัด คือ ผู้สอบสามารถลอกคำตอบกันได้ง่ายกว่าข้อสอบแบบเขียนตอบ ดังนั้นจึงมีเทคนิคที่ครูหรือผู้สร้างข้อสอบใช้ในการป้องกันการลอกข้อสอบคือ การเพิ่มจำนวนชุดของข้อสอบ โดยภายในข้อสอบแต่ละชุดจะมีลำดับข้อสอบที่สลับกัน ซึ่งจะช่วยป้องกันการลอกข้อสอบหรือการโกงกันได้ แต่การนำข้อสอบมาสลับลำดับข้อคำถามนั้น ครูหรือผู้สร้างข้อสอบอาจไม่ได้คำนึงถึงลำดับความยากง่ายของข้อสอบ หรือลำดับของเนื้อหา ซึ่งส่งผลกระทบต่อความสามารถที่แท้จริงของผู้สอบได้ ซึ่งจากผลการวิจัยในครั้งนี้พบว่า การเรียงลำดับข้อสอบทั้ง 3 รูปแบบ ไม่มีผลต่อการทวนคำตอบของผู้สอบในทุกระดับความสามารถ ดังนั้นในการวิจัยครั้งต่อไปควรศึกษาการเรียงลำดับข้อสอบทั้ง 3 รูปแบบร่วมกับการตรวจสอบการคัดลอกคำตอบ (Detecting powers of answer copying) เพื่อตรวจสอบว่ารูปแบบการเรียงลำดับข้อสอบนั้นสามารถป้องกันการลอกข้อสอบได้มากน้อยเพียงใด

3. การวิจัยในครั้งนี้เป็นการเปรียบเทียบผลที่ได้จากการทวนคำตอบ ในข้อสอบแบบหลายตัวเลือก ที่มีการเรียงลำดับข้อสอบ ระยะเวลาในการสอบและระดับความสามารถของผู้สอบที่แตกต่างกันซึ่งเป็นข้อสอบในระดับ Low-stakes test หรือ ข้อสอบที่มีความสำคัญหรือความสำคัญต่ำ ที่ผู้สอบน่าจะมีแรงจูงใจในการสอบหรือมีความกระตือรือร้นในการสอบต่ำเมื่อเทียบกับ ข้อสอบระดับ High-stakes test ดังนั้นในการวิจัยครั้งต่อไป ควรศึกษาผลที่ได้จากการทวนคำตอบกับข้อสอบระดับ High-stakes test หรือ ข้อสอบที่มีความสำคัญสูงหรือมีความเสี่ยงสูง ซึ่งน่าจะให้สารสนเทศในรายละเอียดที่เพิ่มขึ้นและต่อยอดองค์ความรู้เดิมมากยิ่งขึ้นต่อไป

รายการอ้างอิง

ภาษาไทย

- กระทรวงศึกษาธิการ. (2556). *ระบบจัดเก็บข้อมูลนักเรียนรายบุคคล*. สืบค้นวันที่ 15 กันยายน 2556, จาก <http://portal.bopp-obec.info/obec56>
- กานดา ทองวัฒนะ. (2521). *การเปลี่ยนคำตอบในการทำข้อสอบชนิดเลือกตอบ*. วิทยานิพนธ์ระดับปริญญา มหาบัณฑิต สาขาวิชาการวัดและประเมินผลการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.
- ขวัญใจ อุดรธารณ์. (2528). *การเปลี่ยนคำตอบในการสอบซ้ำของแบบสอบเลือกตอบวิชา คณิตศาสตร์ ชั้นมัธยมศึกษาตอนต้น*. วิทยานิพนธ์ระดับปริญญา มหาบัณฑิต สาขาวิชาการวัดและประเมินผลการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.
- ชลินทร์ ศิลปสมศักดิ์. (2544). *การเปรียบเทียบความเชื่อมั่นของแบบทดสอบเลือกตอบถูกผิดที่มีการตรวจให้คะแนนตามกฎชิลลิ่งโดยมีการจัดเรียงข้อสอบและและกำหนดจุดชิลลิ่งต่างกัน*. วิทยานิพนธ์ระดับปริญญา มหาบัณฑิต สาขาวิชาการวัดผลการศึกษา บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ.
- โชติกา ภาชีผล. (2555). *การวัดและการประเมินผลการศึกษา*. กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

- ประณาท เทียนศรี. (2546). ปัจจัยและผลการเปลี่ยนคำตอบข้อสอบปรนัยของนักเรียนชั้นประถมศึกษาปีที่ 5 โรงเรียนสาธิตจุฬาลงกรณ์มหาวิทยาลัย. *วารสารครุศาสตร์*, 31, 10-23.
- พิมพ์สิริ เจริญเศรษฐ์. (2549). การเปรียบเทียบค่าความสามารถ ลักษณะของการเปลี่ยนคำตอบ และเวลาที่ใช้ในการทำข้อสอบของผู้สอบ ในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ เมื่อเงื่อนไขการทดสอบและระดับความสามารถของผู้สอบแตกต่างกัน. วิทยานิพนธ์ระดับปริญญาโท สาขาวิชาการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.
- ศิริวิมล สุขสมบูรณ์. (2541). ผลการวิเคราะห์แบบแผนการตอบข้อสอบของแบบทดสอบเลือกตอบที่มีรูปแบบการจัดเรียงข้อความต่างกัน. วิทยานิพนธ์ระดับปริญญาโท สาขาวิชาการศึกษา บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร.

ภาษาอังกฤษ

- Aamodt, M. G., & McShane, T. (1992). A meta-analytic investigation of the effect of various test item characteristics on test scores. *Academic Journal*, 21(2), 151-160.
- Ahmann, J. S., & Glock, M. D. (1971). *Measuring and evaluating educational achievement*. Boston: Allyn and Bacon.
- Al-Hamly, M., & Coombe, C. (2005). To change or not to change: Investigating the value of MCQ answer changing for Gulf Arab students. *Language Testing*, 22(4), 509-531.
- Bridgeman, B. (2012). A simple answer to a simple question on changing answers. *Journal of Educational Measurement*, 49, 467-468.
- Cohen, J. (1977). Statistical Power Analysis and Research Results. *American Educational Research Journal*, 10, 225-229.
- Di Milia, L. (2007). Benefiting from multiple-choice exams: The positive impact of answer switching. *Educational Psychology*, 27(5), 607-615
- Ferguson, K. J., Kreiter, C. D., Peterson, M. W., Rowat, J. A., & Elliott, S. T. (2002). Is that your final answer? Relationship of changed answers to overall performance on a computer-based medical school course examination. *Teach Learn Med*, 14(1), 20-23.
- Foot, R., & Belinky, C. (1972). It pays to switch? Consequences of changing answers on multiple-choice examination. *Psychological Reports*, 31, 667-673.
- Haladyna, T. M., & Downing, S. M. (1989). A taxonomy of multiple choice item-writing rules. *Applied Measurement in Education*, Vol. 1, pp. 37-50.
- Kastner, M., & Stangl, B. (2011). Multiple choice and constructed response tests: Do test format and scoring matter? *Procedia Social and Behavioral Sciences*, 12, 263-273.
- Mueller, D. J., & Wasser, V. (1977). Implication of changing answers on objective test items. *Journal of Educational Measurement*, 14(1), 9-13.

- Reile, & Briggs. (1952). Should students change their initial answers on objective-type tests? More evidence regarding an old problem. *Journal of Educational Psychology*, 42, 110-115.
- Van der Linden, W. J., Jeon, M., & Ferrara, S. (2011). A paradox in the study of the benefits of test-item review. *Journal of Educational Measurement*, 48(4), 380-398.
- Vispoel, W. P. (1998). Reviewing and changing answers on computer adaptive and self-adaptive vocabulary tests. *Journal of Educational Measurement*, 35, 328-447.