ผลของการให้คะแนนที่มีต่อประสิทธิภาพการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์

The Effects of Scoring Methods on Efficiency of Computerized Adaptive Testing

ภาณุวัชร ปุรณะศิริ [1]**/ สมศักดิ์ ลิลา** [2]**/ สมพงษ์ ปั้นหุ่น** [3]

Panuwat Puranasiri / Somsak Lila / Sompong Panhoon

[1] สาขาวิชาวิจัย วัดผล และสถิติการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา

Program in Educational Research Measurement and Statistics, Faculty of Education, Burapha University

[2-3] ภาควิชาวิจัยและจิตวิทยาประยุกต์ คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา

Department of Research and Applied Psychology, Faculty of Education, Burapha University

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาผลของการให้คะแนนที่มีต่อคุณภาพการวัดด้านความตรง ความเที่ยง ความคลาดเคลื่อนมาตรฐานในการวัด และเปรียบเทียบผลของการให้คะแนนที่มีต่อประสิทธิภาพการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ในด้านจำนวนข้อสอบ และค่าฟังก์ชั่นสารสนเทศของชุดข้อสอบ เมื่อพิจารณาตามเกณฑ์การคัดเลือกข้อสอบ การประมาณค่าความสามารถผู้สอบ การยุติการทดสอบ และความสามารถของผู้สอบ โดยระยะแรกเป็นการสร้างคลังข้อสอบวิชาคณิตศาสตร์ ระดับชั้นมัธยมศึกษาตอนปลาย วิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ กลุ่มตัวอย่าง 3,330 คน ได้ข้อสอบ 230 ข้อ ยังไม่ได้ปรับสเกล การวัดนั้นมีลักษณะ ใกล้เคียงกัน ซึ่งสังเกตได้จากค่าเฉลี่ยของค่าพารามิเตอร์ของข้อสอบที่วิเคราะห์ได้ มีค่าอำนาจจำแนก 0.73-0.95 และค่าเบี่ยงเบนมาตรฐาน 0.06-0.24 อยู่ในเกณฑ์ดี มีค่าความยากง่าย 0.79-1.09 และค่าเบี่ยงเบนมาตรฐาน 0.29-1.27 อยู่ในเกณฑ์ค่อนข้างยาก มีค่าการเดา 0.11-0.16 และค่าเบี่ยงเบนมาตรฐาน 0.02-0.04 อยู่ในเกณฑ์ดี ทำการตรวจสอบความเป็นมิติเดียวโดยการวิเคราะห์องค์ประกอบ พบว่าค่าไอเกนตัวประกอบที่ 1 สูงกว่าค่าไอเกนตัวประกอบอื่นๆ ที่เหลือที่มีค่าไอเกนใกล้เคียงกัน แสดงว่า ข้อสอบที่ได้มีความเป็นมิติเดียว ระยะที่สองศึกษาผลของการให้คะแนนที่มีต่อประสิทธิภาพการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ วิเคราะห์ข้อมูลกลุ่มตัวอย่าง 540 คน โดยใช้ค่าเฉลี่ยเลขคณิต ค่าเบี่ยงเบนมาตรฐาน ผลการวิจัย พบว่า คุณภาพของการวัดจากผลของการให้คะแนนแบบ Multiple-Response Method (MR) แบบ Multiple True-False Method (MTF) และแบบ Omit Multiple True-False Method (OMTF) มีค่าความตรง คือ 0.7202, 0.7233, 0.7239 ตามลำดับ ค่าความเที่ยง คือ 0.7716, 0.7750, 0.7757 ตามลำดับ และความคลาดเคลื่อนมาตรฐานในการวัด คือ 0.2326, 0.4609, 0.2305 ตามลำดับ สำหรับผลของการให้คะแนนแบบ MR แบบ MTF และแบบ OMTF เมื่อพิจารณาตามเกณฑ์การคัดเลือกข้อสอบ การประมาณค่าความสามารถผู้สอบ การยุติการทดสอบ และความสามารถของผู้สอบ มีผลต่อจำนวนข้อสอบและค่าฟังก์ชั่นสารสนเทศของชุดข้อสอบ

คำสำคัญ: ผลของการให้คะแนน การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์

Abstract

The purposes of this research were to study effects of scoring methods on the quality measure of validity, reliability, and standard error of measurement and to compare results of effects of scoring methods on efficiency of computerized adaptive testing to investigate on number of items and test functional information by considering the composition of the main testing, item selection criteria, ability estimation procedure, termination criteria, examinee's different abilities. The first phase was developing test item bank on Mathematics for an upper secondary school level. The test was analyzed to get the qualities of items by using Item Response Theory. The sample of 3,330 examinee. A total of 230 test items were created that have not adapted the scale, The measurements were similar, as observed by the mean of the analyzed test parameters, having the discrimination index were 0.73-0.95, the standard deviation were 0.06-0.23, which is considered good. The difficulty of test items were 0.79-1.09, the standard deviation were 0.29-1.27. The guessing values were 0.11-0.16, the standard deviation were 0.02-0.04, which is considered good. Having checked the unidimensional of the test by using factor analysis, it revealed that the first value was higher than other values, with similar values all together. It is assumed that the test is unidimensional. The second phase was studying the effects of scoring methods by the five independent variables including selection criteria, ability estimation procedure, termination criteria and examinee's ability on efficiency of computerized adaptive testing to investigate on number of items and test functional information. The data obtained from 540 samples were analyzed by using means, standard deviation, and Analysis of Variance. The results indicated that the quality of the measurement results scoring of Multiple-Response Method (MR), Multiple True-False Method (MTF) and Omit Multiple True-False Method (OMTF) revealed validity of 0.7202, 0.7233, 0.7239, respectively. The reliability values were 0.7716, 0.7750, 0.7757, respectively. The standard error of measurement values were 0.2326, 0.4609, 0.2305, respectively. The Effects of MR, MTF and OMTF scoring after item considering selection criteria, ability estimation procedure, termination criteria examinee's ability, number of items, and functional test information.

Keywords: Effects of Scoring Methods, Computerized Adaptive Testing

บทนำ

การทดสอบทางการศึกษาเป็นกระบวนการวัดทางจิตวิทยาเพื่อค้นหาและอธิบายพฤติกรรมจากการวัดคุณลักษณะภายในของมนุษย์ ดังคำกล่าวที่ว่า Test to discover and develop human talent แปลว่า ทดสอบเพื่อค้นและพัฒนาสมรรถภาพของมนุษย์ (ชวาล แพรัตกุล, 2552 : 199) ในการวัดคุณลักษณะที่แฝงอยู่ภายในตัวบุคคลไม่สามารถสังเกตได้โดยตรงต้องอาศัยทฤษฎีการทดสอบและระบบการทดสอบที่มีประสิทธิภาพ การนำเทคโนโลยีมาใช้บนพื้นฐานของทฤษฎีการวัดแนวใหม่ จะช่วยเสริมประสิทธิภาพของสิ่งเร้าคือ ข้อสอบ พัฒนาคลังข้อสอบ การจัดแบบสอบ การตรวจข้อสอบ การวิเคราะห์ข้อสอบ การแปลผลการสอบ และการรายงานผล ทำให้ผลการสอบมีความแม่นตรง กระชับ รัดกุม และยืดหยุ่น แนวโน้มของการพัฒนาเครื่องมือทางเทคโนโลยีคอมพิวเตอร์มีบทบาทมากขึ้นเพราะสามารถหาได้ง่าย ราคาถูก มีประสิทธิภาพและยังช่วยเพิ่มคุณภาพและประสิทธิภาพของการวัดผลการเรียนรู้ของบุคคลได้อย่างรวดเร็ว ถูกต้อง ทำให้กระบวนการวัดผลมีความคล่องตัว (สมศักดิ์ ลิลา, 2539 : 52-53) ในยุคแรกของการทดสอบแบบนี้เรียกว่า การทดสอบแบบเทเลอร์ (Tailored testing) คำนี้ใช้เป็นครั้งแรก โดยเทิร์นบูล Turnbull (1951) (Cited in Lord, 1980) ต่อมา ไวส์ (Weiss, 1974) ได้เสนอให้ใช้คำว่า Adaptive test (อ้างถึง

ใน ศิริชัย กาญจนวาสี, 2550 : 193) ซึ่งเป็นกระบวนการประมาณค่าความสามารถของผู้สอบอย่างต่อเนื่อง และจัดข้อสอบที่เหมาะสมกับระดับความสามารถของผู้สอบ เรียกวิธีการนี้ว่า Computerized adaptive testing ใช้ตัวย่อว่า CAT (บุญชม ศรีสะอาด, 2540 : 73) การทดสอบจึงเริ่มพลิกโฉมหน้าใหม่ด้วยการใช้คอมพิวเตอร์สำหรับการทดสอบบนพื้นฐานของทฤษฎีการตอบสนองข้อสอบที่มีประสิทธิภาพ และถูกต้องแม่นยำยิ่งขึ้น การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์มีจุดเด่น คือ ทำให้การจัดการสอบสะดวก สนองตอบต่อการทดสอบรายบุคคล การเลือกข้อสอบที่เหมาะสมกับความสามารถของผู้สอบ การทดสอบมีสภาพสอดคล้องกับธรรมชาติของผู้สอบมากขึ้น ความวิตกกังวลในการสอบลดลง และยังสามารถประมวลผลได้อย่างรวดเร็ว รายงานผลได้ทันที และมีความถูกต้องแม่นยำมากกว่าการทดสอบแบบประเพณีนิยม (ศิริชัย กาญจนวาสี, 2550 : 220)

การวัดสมรรถภาพของบุคคลโดยใช้แบบทดสอบนั้นได้รับความนิยมเป็นอย่างมาก เพราะมีความเป็นปรนัยในการตรวจให้คะแนน ใช้เวลารวดเร็วในการตรวจให้คะแนน ผลการวิเคราะห์ข้อสอบสามารถนำมาวินิจฉัยปัญหาการเรียนรู้ได้ และสามารถใช้คอมพิวเตอร์ตรวจให้คะแนนได้อีกด้วย แต่วิธีการตอบและการให้คะแนนที่เปิดโอกาสให้ผู้สอบมีการตอบถูกโดยการเดานั้น จะส่งผลต่อความคลาดเคลื่อนในการวัดผล และการให้คะแนนบางวิธียังไม่สามารถให้สารสนเทศที่เพียงพอที่จะจำแนกผู้สอบระดับต่างๆ ของความรู้ได้ ระหว่างผู้ที่มีความรู้สมบูรณ์ (Full knowledge) ผู้ที่มีความรู้บางส่วน (Partial knowledge) และผู้ที่ไม่มีความรู้ (Absence of knowledge) จึงได้มีการศึกษาวิธีการแก้ปัญหา เพื่อลดโอกาสในการเดาสุ่ม และเพิ่มสารสนเทศเกี่ยวกับการตอบข้อสอบให้ได้มากที่สุด โดยปรับปรุงวิธีการให้คะแนนซึ่งเป็นแหล่งอ้างอิงสำคัญของการวัดผลจิตวิทยา การให้คะแนนที่เหมาะสม สามารถบอกถึงตำแหน่งจริงของความรู้ของผู้สอบ ทำให้คุณภาพของการวัดผลเกี่ยวกับความเที่ยงและความตรงสูง ทำให้คะแนนมีความเชื่อมั่นมากขึ้น (Frary, 1980 : 79-90)

ปัจจุบันการทดสอบ (Testing) โดยใช้ข้อสอบ (Item) กระตุ้นการตอบสนองต่อผู้สอบแล้วนับหน่วยความสามารถเป็นคะแนนรายข้อนั้น จะรวมความคลาดเคลื่อนในการวัด (Measurement error) อยู่ด้วย ฉะนั้นวิธีการคิดคะแนนความสามารถที่เหมาะสมต้องสอดคล้องกับโมเดลการคิดของผู้สอบ จึงจะทำให้การสรุปเกี่ยวกับความสามารถของผู้สอบถูกต้อง แม่นยำ มีความคลาดเคลื่อนต่ำ คะแนนที่วัดได้จึงจะสอดคล้องกับความสามารถที่แท้จริงมากที่สุด (Lord, 1980 : 3) ผลของการให้คะแนน เป็นส่วนสำคัญของการวัดที่ส่งผลให้กระบวนการทดสอบมีประสิทธิภาพสนองตอบต่อผู้สอบในเชิงจิตวิทยา ผู้สอบที่มีความสามารถสูงควรจะได้รับข้อสอบที่ยาก ส่วนผู้สอบที่มีความสามารถต่ำก็ควรจะได้รับข้อสอบที่ง่าย (Lord & Novick, 1968 : 359) ชอบปิน (Choppin, 1971 cited in Leclercq, 1983 : 161-287) ได้เสนอว่า กระบวนการวัดผลที่ให้ผู้สอบได้แสดงความรู้บางส่วนนั้นสอดคล้องกับโมเดลในการคิดของผู้สอบ ซึ่งการทดสอบและการให้คะแนนมีความสัมพันธ์ใกล้ชิดกับโมเดลทางความรู้ (Epistemological model) โดยได้เสนอแนวคิดในเรื่องการแสดงออก ซึ่งความรู้ของผู้ตอบแบบสอบเลือกตอบเป็นลักษณะต่อเนื่อง ได้แก่ โมเดลที่สองเสนอว่า “เมื่อผู้ตอบรู้คำตอบ เขาจะเลือกตัวเลือกที่เป็นคำตอบที่ถูกต้องในข้อนั้น แต่เมื่อผู้ตอบไม่รู้คำตอบ เขาจะตัดตัวเลือกที่เขารู้ว่าไม่ถูกต้องออกและเขาจะเลือกอย่างสุ่มในระหว่างตัวเลือกที่เหลือ” ส่วนโมเดลที่สามเสนอว่า “ขั้นความรู้ของผู้ตอบไม่สามารถแบ่งได้ เป็นรู้คำตอบกับไม่รู้คำตอบเท่านั้น แต่มีความต่อเนื่องของขั้นความรู้จากขั้นที่ไม่รู้อย่างสมบูรณ์ไปสู่ขั้นที่รู้อย่างสมบูรณ์ ซึ่งในช่วงระหว่างขั้น ของความรู้นี้ เรียกว่า ความรู้บางส่วน ซึ่งเมื่อผู้ตอบเผชิญกับปัญหาที่เป็นตัวเลือกต่าง ๆ เขาจะตอบสนองต่อตัวเลือกนั้นแต่ละตัว ตามค่าความน่าจะเป็นของการที่จะเป็นคำตอบที่ถูกต้อง แต่ด้วยคำชี้แจงในการตอบที่ให้เขาเลือกเพียงตัวเลือกเดียว เขาจึงต้องเลือกตัวเลือกที่มีความน่าจะเป็นว่าจะถูกต้องมากที่สุด” คูมบ์และคณะ (Coombs, Milholland, & Womer, 1956 : 13-17) จึงได้เปลี่ยนแปลงวิธีการตอบข้อสอบ โดยใช้ข้อสอบเลือกตอบ (Multiple True-False: MTF) ซึ่งผู้สอบจะต้องพิจารณาคำตอบทุกตัวเลือก การตอบในลักษณะนี้ เป็นการวัดความคิดรวบยอด โดยผู้สอบจะต้องแยกแยะสิ่งที่เป็นความคิดรวบยอดกับสิ่งที่ไม่เป็นออกจากกัน จึงจะถือได้ว่ารู้จริง

การบูรณาการศาสตร์แห่งการวัดผลที่มีวิธีการให้คะแนนที่เหมาะสม กับเทคโนโลยีการทดสอบด้วย

คอมพิวเตอร์เข้าด้วยกัน จะทำให้กระบวนการทดสอบมีความคล่องตัว สะดวก มีประสิทธิภาพสูง โดยมีองค์ประกอบที่สำคัญ ได้แก่ การคัดเลือกข้อสอบจากคลังข้อสอบ (Item selection rule) การประมาณค่าความสามารถของผู้สอบ (Ability estimation procedure) และการยุติการทดสอบ (Termination criteria) (ศิริชัย กาญจนวาสี, 2550 : 201-206) และในปัจจุบัน การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ใช้ประมาณค่าความสามารถของผู้สอบ จากผลการให้คะแนนแบบสองค่าหรือถูก 1 ผิด 0 ที่อาจเกิดความคลาดเคลื่อนในการวัดได้ง่าย จึงควรที่จะมีการศึกษาผลของการให้คะแนนแบบหลายค่า โดยใช้ข้อสอบเลือกตอบ (Multiple True-False) ที่มีการให้คะแนนแบบ Multiple-Response Method (MR) แบบ Multiple True-False Method (MTF) และแบบ Omit Multiple True-False Method (OMTF) ว่าจะส่งผลต่อประสิทธิภาพของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ในด้านจำนวนข้อสอบและค่าฟังก์ชั่นสารสนเทศของชุดข้อสอบหรือไม่ มากหรือน้อยเพียงไร เพื่ออธิบายผลของการให้คะแนนที่ใกล้เคียงกับความสามารถจริงให้มากที่สุด

วัตถุประสงค์ของการวิจัย

1. เพื่อศึกษาผลของการให้คะแนนที่มีต่อคุณภาพการวัดผลด้านความตรง ความเที่ยง และความคลาดเคลื่อนมาตรฐานในการวัด

2. เพื่อเปรียบเทียบผลของการให้คะแนนที่มีต่อประสิทธิภาพการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ด้านจำนวนข้อสอบ และค่าฟังก์ชั่นสารสนเทศของชุดข้อสอบ เมื่อกำหนดให้การคัดเลือกข้อสอบ การประมาณค่าความสามารถผู้สอบ การยุติการทดสอบ และความสามารถของผู้สอบต่างกัน

กรอบแนวคิดของการวิจัย

ผลของการให้คะแนน 3 แบบ ได้แก่ 1) Multiple-Response Method (MR) เป็นแบบของการให้คะแนนแก่ความรู้ที่ถูกต้องสมบูรณ์ คือ ผู้สอบจะต้องตอบได้ตรงกับสภาพความเป็นจริงหรือตรงตามเฉลยทุกตัวเลือก จึงจะได้ 1 คะแนน แต่ถ้าไม่ตรงกับสภาพความเป็นจริงแม้เพียงตัวเลือกเดียวก็จะได้ 0 คะแนน วิธีนี้บ่งชี้ว่าผู้สอบมีความรู้ความสามารถในเรื่องนั้นๆ ถูกต้องสมบูรณ์หรือไม่ ถ้าได้ 0 คะแนน ก็แสดงว่า ยังมีบางประเด็นที่บกพร่องอยู่ ดังนั้นคะแนนที่เป็นไปได้ของข้อสอบที่ตรวจให้คะแนนแบบนี้จึงมีค่า 2 ค่า ได้แก่ 0 และ 1 คะแนน 2) Multiple True-False Method (MTF) เป็นแบบของการตรวจให้คะแนนทุกตัวเลือก ถ้าตัวเลือกใดนักเรียนกาได้ถูกต้องจะได้ 1 คะแนน ถ้ากาผิดจะได้คะแนน 0 ดังนั้น ถ้าข้อสอบมี 4 ตัวเลือก ก็จะได้คะแนนอยู่ระหว่าง 0-4 คะแนน 3) Omit Multiple True-False Method (OMTF) เป็นแบบของการตรวจให้คะแนนทุกตัวเลือก โดยคำนึงถึงตัวเลือกที่ผู้สอบเว้นไว้จะไม่คิดคะแนน เพื่อให้เกิดความยุติธรรมแก่ผู้สอบที่มีความรู้เต็ม (Full knowledge) ที่ควรได้คะแนนสูงสุด ให้มีค่าเท่ากับ 4 หรือผู้ที่ไม่มีความรู้ (Null knowledge) ที่ควรได้คะแนนต่ำสุด ให้มีค่าเท่ากับ 0 ดังนั้น ถ้าข้อสอบมี 4 ตัวเลือก ก็จะได้คะแนนอยู่ระหว่าง 0-4 คะแนน ทั้ง 3 แบบเป็นวิธีการที่มีความสะดวกในการใช้ สามารถที่จะพัฒนาวิธีการตอบข้อสอบและการให้คะแนนในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ได้อย่างยุติธรรม ซึ่งผู้วิจัยได้ทำการศึกษาแบ่งเป็น 2 ส่วน ได้แก่ การศึกษาคุณภาพของการวัดในด้านความตรง ความเที่ยง และความคลาดเคลื่อนมาตรฐานในการวัด โดยการหาความสัมพันธ์จากผลการให้คะแนนการสอบกับผลคะแนนจากการทดสอบระดับชาติขั้นพื้นฐาน (O-NET) ของนักเรียนระดับชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2557 ทำการศึกษาความตรงเชิงทฤษฎีของการให้คะแนนจากการสอบข้อสอบที่ผู้วิจัยสร้างขึ้น ที่ผ่านการวิเคราะห์คุณภาพของข้อสอบ ตามหลักการของทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์ มีเกณฑ์การคัดเลือกข้อสอบในการจัดทำคลังข้อสอบ โดยข้อสอบแต่ละข้อควรมีค่าอำนาจจำแนกสูง

($a > 0.80$) มีค่าความยากที่ครอบคลุมระดับช่วงต่างๆ อย่างเหมาะสม [$-2.0<\theta<+2.0$] ค่าสัมประสิทธิ์การเดาต่ำ ($c < 0.30$) และควรมีจำนวนข้อสอบเท่าๆ กันในแต่ละระดับความยาก (Urry, 1977)

ในส่วนของการศึกษาประสิทธิภาพการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์นั้น ผู้วิจัยทำการศึกษาจำนวนข้อสอบ และค่าฟังก์ชั่นสารสนเทศชุดข้อสอบ

พิจารณาตามองค์ประกอบสำคัญของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ได้แก่ การคัดเลือกข้อสอบ การประมาณค่าความสามารถของผู้สอบ การยุติการทดสอบ และความสามารถของผู้สอบ

วิธีดำเนินการวิจัย

1. ประชากรและกลุ่มตัวอย่าง เป็นนักเรียนระดับชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2557 โรงเรียนมัธยมศึกษาจังหวัดบุรีรัมย์ จำนวน 9,789 คน กลุ่มตัวอย่างได้มาโดยการสุ่มแบบหลายขั้นตอน (Multi stage random sampling) สำหรับใช้ในการสร้างคลังข้อสอบ จำนวน 4,096 คน สำหรับใช้วิเคราะห์ค่าคุณภาพของการวัด กำหนดขนาดกลุ่มตัวอย่างโดยใช้ตาราง Krejcie and Morgan (บุญชม ศรีสะอาด, 2553 : 43) ได้จำนวนตัวอย่าง 370 คน และใช้วิธีการสุ่มอย่างง่ายเพื่อให้ได้จำนวนกลุ่มตัวอย่างเข้าสอบตามชุดข้อสอบ จำนวน 9 ฉบับ รวมจำนวนทั้งสิ้น 3,330 คน และสำหรับวิเคราะห์ค่าประสิทธิภาพการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ เป็นนักเรียนโรงเรียนบุรีรัมย์พิทยาคม จังหวัดบุรีรัมย์ จำนวน 16 ห้องเรียน จำนวน 650 คน สุ่มนักเรียนโดยแบ่งตามความสามารถทางคณิตศาสตร์ ได้แก่ ความสามารถสูง ปานกลาง และต่ำ ผู้วิจัยต้องการกลุ่มตัวอย่าง จำนวน 540 คน เพื่อใช้เป็นกลุ่มผู้เข้าสอบโดยใช้โปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ จำนวน 6 โปรแกรมๆ ละ 90 คน

2. เครื่องมือที่ใช้ในการวิจัย เป็นข้อสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ ระดับ ชั้นมัธยมศึกษาตอนปลาย ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 เป็นชุดข้อสอบเลือกตอบที่ผู้วิจัยสร้างขึ้น จำนวน 230 ข้อ ที่ยังไม่ได้ปรับสเกล นำไปทดสอบกับกลุ่มตัวอย่างจำนวน 4,096 คน หาค่าคุณภาพตามทฤษฎีการทดสอบแบบดั้งเดิม (CTT) ได้ค่าความยาก ระหว่าง 0.17-0.96 และค่าอำนาจจำแนก ระหว่าง 0.11-0.56 พิจารณาข้อสอบที่มีคุณภาพได้จำนวน 190 ข้อ จัดพิมพ์เป็นชุดข้อสอบ จำนวน 9 ฉบับ ฉบับละ 30 ข้อ (แต่ละฉบับมีข้อสอบต่างจากฉบับอื่นอยู่ 20 ข้อ และมีข้อสอบที่เหมือนกันอยู่ 10 ข้อในทุกฉบับซึ่งใช้เป็นข้อสอบร่วม) นำไปทดสอบกับกลุ่มตัวอย่าง จำนวน 3,330 คน หาค่าคุณภาพโดยใช้การวิเคราะห์พารามิเตอร์ข้อสอบตามทฤษฎีการตอบสนองข้อสอบ (IRT)

3. การดำเนินการวิจัยแบ่งเป็น 2 ระยะ ได้แก่ ระยะแรกดำเนินการสร้างคลังข้อสอบวิชาคณิตศาสตร์ระดับชั้นมัธยมศึกษาตอนปลาย ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 และระยะที่สองเป็นการศึกษาคุณภาพของการวัดและประสิทธิภาพของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ในด้านจำนวนข้อสอบ และค่าฟังก์ชั่นสารสนเทศของชุดข้อสอบ ที่เกิดจากผลของการให้คะแนนพิจารณาตามองค์ประกอบสำคัญ ได้แก่ การคัดเลือกข้อสอบ การประมาณค่าความสามารถผู้สอบ การยุติการทดสอบ และความสามารถของผู้สอบ ดังนี้

ขั้นตอนที่ 1 การวิเคราะห์ข้อมูลเพื่อหาค่าพารามิเตอร์ของข้อสอบและนำเข้าคลังข้อสอบ ศึกษาตรวจสอบคุณภาพของข้อสอบในด้านความตรง ความเที่ยง และความคลาดเคลื่อนมาตรฐานในการวัดของการให้คะแนนแบบ MR แบบ MTF และแบบ OMTF

ขั้นตอนที่ 2 พัฒนาโปรแกรมการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ เป็นการออกแบบโปรแกรมคอมพิวเตอร์ และเขียนโปรแกรมด้วยภาษา ASP.NET โดยนำข้อสอบจากคลังข้อสอบในขั้นตอนที่ 1 มาเป็นฐานข้อมูลสำหรับใช้ในการพัฒนาโปรแกรม จากนั้นให้ผู้เชี่ยวชาญพิจารณาตรวจสอบความถูกต้องเหมาะสมของโปรแกรม ปรับปรุงตามคำแนะนำของผู้เชี่ยวชาญแล้วทำการทดลองใช้โปรแกรมการทดสอบ โดยนำโปรแกรมย่อยทั้ง 6 โปรแกรมไปให้นักเรียนทดลองใช้งาน ปรับปรุงตามคำแนะนำของนักเรียนและจัดทำเป็นชุดโปรแกรมสำหรับการทดสอบ ดังนี้

โปรแกรมย่อย	การคัดเลือกข้อสอบ	การประมาณค่า
1	$[Match\ b_i\ to\ \hat{\theta}]$	Conditional maximum likelihood (MLE)
2	$[Match\ b_i\ to\ \hat{\theta}]$	Bayesian Updating (Bay)
3	$[Match\ Q_{max}\ to\ \hat{\theta}]$	Conditional maximum likelihood (MLE)
4	$[Match\ Q_{max}\ to\ \hat{\theta}]$	Bayesian Updating (Bay)
5	I_{max}	Conditional maximum likelihood (MLE)
6	I_{max}	Bayesian Updating (Bay)

ขั้นตอนที่ 3 ดำเนินการทดสอบกับนักเรียนกลุ่มตัวอย่างโดยใช้โปรแกรมคอมพิวเตอร์ เพื่อศึกษาประสิทธิภาพการทดสอบ ในด้านจำนวนข้อสอบและค่าฟังก์ชั่นสารสนเทศของชุดข้อสอบ พิจารณาตามเกณฑ์การคัดเลือกข้อสอบ (Item selection rule) การประมาณค่าความสามารถของผู้สอบ (Ability estimation procedure) เกณฑ์ยุติการทดสอบ (Termination criteria) และ ความสามารถของผู้สอบ (Examinee's ability)

การวิเคราะห์ข้อมูล

1. ค่าสถิติพื้นฐาน โดยการหาค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของค่าพารามิเตอร์ข้อสอบ การประมาณค่าพารามิเตอร์ของข้อสอบ โดยวิเคราะห์ข้อสอบรายข้อตามรูปแบบโลจิสติก 3 พารามิเตอร์ เพื่อให้ได้ค่าอำนาจจำแนก (a) ค่าความยาก (b) ค่าการเดา (c) ด้วยโปรแกรมสำเร็จรูปไบล็อค (BILOG Version 3.04) ซึ่งความสัมพันธ์เชิงคณิตศาสตร์ของพารามิเตอร์ตามรูปแบบโมเดลโลจิสติก 3 พารามิเตอร์ เขียนได้ดังนี้ (Hambleton, Swaminathan & Roger, 1991 : 17)

$$P_i(\theta) = c_i + (1-c_i)\frac{e^{Da_i(1-b_i)}}{1+e^{Da_i(1-c_i)}};\ i = 1, 2, 3,...,n$$

เมื่อ แทน $P_i(\theta)$ โอกาสการตอบที่ผู้มีความสามารถ (θ) จะตอบข้อสอบถูก

แทน θ ระดับความสามารถของผู้สอบ

แทน a_i ค่าอำนาจจำแนกของข้อสอบข้อ i

แทน b_i ค่าความยากของข้อสอบข้อ i

แทน c_i ค่าการเดาของข้อสอบข้อ i

แทน D Scaling factor มีค่าเท่ากับ 1.7

แทน e ค่าคงที่ มีค่าเท่ากับ 2.71828...

2. วิเคราะห์เพื่อตรวจสอบความเป็นมิติเดียวของแบบสอบ ใช้วิธีการวิเคราะห์องค์ประกอบ (Factor analysis) โดยการวิเคราะห์นี้ใช้ตัวประกอบสำคัญ (Principal component) แล้วหมุนแกนด้วยวิธีแวริแมกซ์ (Varimax) พิจารณาค่าไอเกน (Eigen value) ถ้าหากค่าไอเกนของตัวประกอบที่ 1 ของแบบสอบมีค่าสูงกว่าค่าไอเกนของตัวประกอบที่ 2 อย่างมากและค่าไอเกนของตัวประกอบที่ 2 สูงกว่าค่าไอเกนของตัวประกอบถัดไปเพียงเล็กน้อย จะสรุปได้ว่าแบบสอบทั้งฉบับมีคุณสมบัติการวัดเพียงมิติเดียว (Lord, 1980 : 21)

3. หาค่าคุณภาพของข้อสอบเลือกตอบแบบถูกผิด ได้แก่ ความตรง (Validity) ความเที่ยง (Reliability)

และความคลาดเคลื่อนมาตรฐานในการวัด (Standard error of measurement)

4. หาประสิทธิภาพของการทดสอบ ได้แก่ จำนวนข้อสอบ (Length testing) และฟังก์ชั่นสารสนเทศของชุดข้อสอบ (Test information function) ภายใต้เงื่อนไขของวิธีการให้คะแนนที่ต่างกัน 3 แบบ ได้แก่ แบบ MR แบบ MTF และแบบที่ผู้วิจัยเสนอคือ OMTF ทำการเปรียบเทียบตามองค์ประกอบสำคัญของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ได้แก่ เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ (Item selection rule) การประมาณค่าความสามารถของผู้สอบ (Ability estimation procedure) เกณฑ์ยุติการทดสอบ (Termination criteria) และความสามารถของผู้สอบ (Examinee's ability)

ผลการวิจัย

จากผลการวิเคราะห์ข้อมูล พบว่า การให้คะแนนแบบ OMTF มีค่าความตรงและความเที่ยงสูงสุด คือ 0.7239, 0.7757 รองลงมาได้แก่ แบบ MTF เท่ากับ 0.7202, 0.7750 และแบบ MR เท่ากับ 0.7233, 0.7716 และความคลาดเคลื่อนมาตรฐานในการวัดของการให้คะแนนแบบ OMTF มีค่าต่ำสุด เท่ากับ 0.2305 รองลงมาได้แก่ แบบ MR เท่ากับ 0.2326 และแบบ MTF เท่ากับ 0.4609 สำหรับประสิทธิภาพของการทดสอบ เมื่อพิจารณาตามองค์ประกอบสำคัญของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์จากผลการให้คะแนนทั้ง 3 แบบ พบว่า

1. การคัดเลือกข้อสอบจากคลังข้อสอบในการทดสอบแบบเหมาะด้วยคอมพิวเตอร์ นั้นปรากฏว่า การคัดเลือกข้อสอบด้วยการใช้ค่าสารสนเทศสูงสุดของข้อสอบจะใช้ข้อสอบน้อยกว่าการคัดเลือกข้อสอบด้วยวิธีอื่นๆ และให้ค่าฟังก์ชั่นสารสนเทศสูงสุด จากผลของการให้คะแนนทั้ง 3 แบบ ซึ่งจำนวนข้อสอบที่ใช้การให้คะแนนแบบ MR ที่ใช้เกณฑ์การคัดเลือกข้อสอบข้อต่อไปจากคลังข้อสอบด้วยวิธีการจับคู่ระหว่างค่าความยากง่ายของข้อสอบ (b) สอดคล้องกับค่าความสามารถที่ประมาณได้ของผู้สอบ ($\hat{\theta}$) มีค่าเฉลี่ยของการใช้ข้อสอบสูงสุด คือ 17.59 ค่าเบี่ยงเบนมาตรฐาน 5.78 และวิธีให้คะแนนแบบ OMTF ที่ใช้เกณฑ์การคัดเลือกข้อสอบข้อต่อไปจากคลังข้อสอบด้วยวิธีการใช้ข้อสอบที่มีค่าสารสนเทศสูงสุด (Maximum iteminformation, I_{max}) มีค่าเฉลี่ยการใช้จำนวนข้อสอบต่ำที่สุด คือ 10.34 และค่าเบี่ยงเบนมาตรฐาน 4.76 สำหรับค่าฟังก์ชั่นสารสนเทศของชุดข้อสอบที่ใช้การให้คะแนนแบบ OMTF ที่ใช้เกณฑ์การคัดเลือกข้อสอบข้อต่อไปจากคลังข้อสอบด้วยวิธีการใช้ข้อสอบที่มีค่าสารสนเทศสูงสุด (Maximum item information,) มีค่าเฉลี่ยของฟังก์ชั่นสารสนเทศสูงที่สุด คือ 12.05 ค่าเบี่ยงเบนมาตรฐาน 0.81 และแบบ MR ที่ใช้เกณฑ์การคัดเลือกข้อสอบข้อต่อไปจากคลังข้อสอบด้วยวิธีการจับคู่ระหว่างค่าความยากง่ายของข้อสอบ (b) สอดคล้องกับค่าความสามารถที่ประมาณได้ของผู้สอบ ($\hat{\theta}$) มีค่าเฉลี่ยฟังก์ชั่นสารสนเทศต่ำที่สุด คือ 11.46 ค่าเบี่ยงเบนมาตรฐาน 0.29 และค่าฟังก์ชั่นสารสนเทศของชุดข้อสอบที่ใช้การให้คะแนนแบบ OMTF ที่ใช้เกณฑ์การคัดเลือกข้อสอบข้อต่อไปจากคลังข้อสอบด้วยวิธีการใช้ข้อสอบที่มีค่าสารสนเทศสูงสุด (maximum item information, I_{max}) มีค่าเฉลี่ยของฟังก์ชั่นสารสนเทศสูงที่สุด คือ 12.05 ค่าเบี่ยงเบนมาตรฐาน 0.80 และแบบ MR ที่ใช้เกณฑ์การคัดเลือกข้อสอบข้อต่อไปด้วยวิธีการจับคู่ระหว่างค่าความยากง่ายของข้อสอบ (b) สอดคล้องกับค่าความสามารถที่ประมาณได้ของผู้สอบ ($\hat{\theta}$) มีค่าเฉลี่ยฟังก์ชั่นสารสนเทศต่ำที่สุด คือ 11.46 ค่าเบี่ยงเบนมาตรฐาน 0.29

2. การประมาณค่าความสามารถของผู้สอบในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ทำให้จำนวนข้อสอบที่ใช้ในการทดสอบต่างกัน วิธีของเบส์ใช้ข้อสอบมากกว่าวิธีความเป็นไปได้สูงสุด และให้ค่าฟังก์ชั่นสารสนเทศใกล้เคียงกัน ซึ่งจำนวนข้อสอบที่ใช้การให้คะแนนแบบ MR ที่ใช้วิธีการประมาณค่าความสามารถของผู้สอบด้วยวิธีของเบส์ที่ปรับใหม่ (Bayesian Updating) มีค่าเฉลี่ยจำนวนข้อสอบที่ใช้สอบสูงสุด คือ 16.32 ค่าเบี่ยงเบนมาตรฐาน 8.69 และแบบ MTF ที่ใช้วิธีการประมาณค่าความสามารถของผู้สอบด้วยวิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข (Conditional maximum likelihood) คือ 12.35 ค่าเบี่ยงเบนมาตรฐาน 5.95 สำหรับค่าฟังก์ชั่นสารสนเทศของชุดข้อสอบที่ใช้การให้คะแนนแบบ OMTF ที่ใช้วิธีการประมาณค่าความสามารถของผู้สอบด้วยวิธีของเบส์ที่ปรับใหม่ (Bayesian Updating) มีค่าฟังก์ชั่นสารสนเทศเฉลี่ย

ของชุดข้อสอบสูงที่สุด คือ 11.85 ค่าเบี่ยงเบนมาตรฐาน 0.50

3. การยุติการทดสอบในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ทำให้จำนวนข้อสอบที่ใช้ในการทดสอบต่างกัน ค่าความคลาดเคลื่อนมาตรฐานในการวัดเท่ากับหรือน้อยกว่า 0.30 ใช้ข้อสอบมากกว่าเกณฑ์อื่น ซึ่งจำนวนข้อสอบที่ใช้การให้คะแนนแบบ MTF ที่ใช้เกณฑ์ยุติการทดสอบโดยใช้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่า 0.30 มีค่าเฉลี่ยของจำนวนข้อสอบสูงที่สุด คือ 16.54 ค่าเบี่ยงเบนมาตรฐาน 8.58 และแบบ OMTF ที่ใช้เกณฑ์ยุติการทดสอบโดยใช้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่า 0.50 มีค่าเฉลี่ยของจำนวนข้อสอบต่ำที่สุด คือ 12.07 ค่าเบี่ยงเบนมาตรฐาน 7.36 สำหรับค่าฟังก์ชั่นสารสนเทศของชุดข้อสอบที่ใช้การให้คะแนนแบบ OMTF ที่ใช้เกณฑ์ยุติการทดสอบโดยใช้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่า 0.50 มีค่าฟังก์ชั่นสารสนเทศของชุดข้อสอบเฉลี่ยสูงที่สุด คือ 12.13 ค่าเบี่ยงเบนมาตรฐาน 0.85 และแบบ MR ที่ใช้เกณฑ์ยุติการทดสอบโดยใช้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่า 0.30 มีค่าฟังก์ชั่นสารสนเทศของชุดข้อสอบต่ำที่สุด คือ 11.37 ค่าเบี่ยงเบนมาตรฐาน 0.23

4. ความสามารถทางคณิตศาสตร์ของผู้สอบในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ทำให้จำนวนข้อสอบที่ใช้ในการทดสอบต่างกัน ผู้สอบที่มีความสามารถสูงจะใช้ข้อสอบน้อยกว่าผู้สอบที่มีความสามารถต่ำ ซึ่งจำนวนข้อสอบที่ใช้การให้คะแนนแบบ MR ของผู้สอบที่มีความสามารถทางคณิตศาสตร์ต่ำ มีค่าเฉลี่ยของการใช้ข้อสอบสูงที่สุด คือ 18.02 ค่าเบี่ยงเบนมาตรฐาน 9.37 และแบบ OMTF ของผู้สอบที่มีความสามารถทางคณิตศาสตร์สูงมีค่าเฉลี่ยของการใช้ข้อสอบต่ำที่สุด คือ 11.09 ค่าเบี่ยงเบนมาตรฐาน 5.70 สำหรับค่าฟังก์ชั่นสารสนเทศของชุดข้อสอบที่ใช้การให้คะแนนแบบ OMTF ของผู้สอบที่มีความสามารถทางคณิตศาสตร์สูง มีค่าฟังก์ชั่นสารสนเทศเฉลี่ยสูงที่สุด คือ 12.15 ค่าเบี่ยงเบนมาตรฐาน 0.86 และแบบ MR ของผู้สอบที่มีความสามารถทางคณิตศาสตร์ปานกลาง มีค่าฟังก์ชั่นสารสนเทศของชุดข้อสอบเฉลี่ยต่ำที่สุด คือ 11.63 ค่าเบี่ยงเบนมาตรฐาน 0.44

อภิปรายผลการวิจัย

1. การคัดเลือกข้อสอบจากคลังข้อสอบในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ปรากฏว่า การให้คะแนนแบบ OMTF ที่ใช้เกณฑ์การคัดเลือกข้อสอบที่มีค่าสารสนเทศสูงสุด (Maximum item information, I_{max}) จะใช้จำนวนข้อสอบและให้ค่าฟังก์ชั่นสารสนเทศของชุดข้อสอบสูงที่สุด ทั้งนี้เพราะว่า การคัดเลือกข้อสอบจากคลังข้อสอบต่างกันในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์นั้น จำนวนข้อสอบและค่าฟังก์ชั่นสารสนเทศของชุดข้อสอบก็ต่างกัน สอดคล้องกับผลการวิจัยของรังสรรค์ มณีเล็ก (2540) ที่ว่าข้อสอบที่ถูกเลือกขึ้นมาใช้แต่ละวิธีจะมีสารสนเทศของข้อสอบแตกต่างกันออกไป ซึ่งค่าสารสนเทศของข้อสอบดังกล่าวจะส่งผลต่อค่าความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ ดังนั้นจึงทำให้วิธีการคัดเลือกข้อสอบแต่ละวิธีลู่เข้าหาความคลาดเคลื่อนในการประมาณค่าที่กำหนดเป็นเกณฑ์ยุติการทดสอบเร็วช้าต่างกัน ซึ่งการเร็วช้าต่างกันนี้ก็คือ การใช้ข้อสอบมากน้อยต่างกันในการทดสอบนั่นเอง

2. การประมาณค่าความสามารถของผู้สอบในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ พบว่า เมื่อใช้วิธีการประมาณค่าความสามารถของผู้สอบต่างกันแล้วจำนวนข้อสอบ และค่าฟังก์ชั่นสารสนเทศก็จะต่างกันไปด้วย ซึ่งวิธีของเบส์จะใช้ข้อสอบมากกว่าวิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข ซึ่งต่างจากข้อค้นพบของต่าย เซี่ยงฉี ที่พบว่า วิธีการประมาณค่าความสามารถของเบส์จะใช้ข้อสอบน้อยกว่าวิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข (ต่าย เซี่ยงฉี, 2534: 122) แต่สำหรับในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์เกี่ยวกับจำนวนข้อสอบ และค่าฟังก์ชั่นสารสนเทศ ของชุดข้อสอบต่างกัน สอดคล้องกับผลการวิจัยของกิฟฟอร์ดและสวามินาทาน (Gifford & Swaminathan, 1990: 33) ที่พบว่า ถ้ากลุ่มตัวอย่างขนาดเล็กและข้อสอบน้อยข้อ การประมาณค่าความสามารถของวิธี Joint Bayesian estimate จะมีความถูกต้องมากกว่า Join maximum likelihood และข้อค้นพบของโฮ (Ho, 1989) ที่พบว่า วิธีการประมาณค่าความสามารถวิธีของเบส์มีความเชื่อมั่นและมีประสิทธิภาพสูงกว่าวิธีความเป็นไปได้สูงสุด ซึ่งมักไม่ค่อย

คงเส้นคงวานัก เมื่อใช้วิธีการประมาณค่าความสามารถของผู้สอบต่างกัน

3. การยุติการทดสอบในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ จำนวนข้อสอบ ที่ใช้การให้คะแนนแบบ MTF ที่ใช้เกณฑ์ยุติการทดสอบโดยใช้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่า 0.30 มีค่าเฉลี่ยของจำนวนข้อสอบสูงที่สุด คือ 16.54 ค่าเบี่ยงเบนมาตรฐาน 8.58 และแบบ OMTF ที่ใช้เกณฑ์ยุติการทดสอบโดยใช้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่า 0.50 มีค่าเฉลี่ยของจำนวนข้อสอบต่ำที่สุด คือ 12.07 ค่าเบี่ยงเบนมาตรฐาน 7.36 และค่าฟังก์ชั่นสารสนเทศของชุดข้อสอบที่ใช้การให้คะแนนแบบ OMTF ที่ใช้เกณฑ์ยุติการทดสอบโดยใช้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่า 0.50 มีค่าฟังก์ชั่นสารสนเทศของชุดข้อสอบเฉลี่ยสูงที่สุด คือ 12.13 ค่าเบี่ยงเบนมาตรฐาน 0.85 และแบบ MR ที่ใช้เกณฑ์ยุติการทดสอบโดยใช้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่า 0.30 มีค่าฟังก์ชั่นสารสนเทศของชุดข้อสอบต่ำที่สุด คือ 11.37 ค่าเบี่ยงเบนมาตรฐาน 0.23 ดังนั้น การยุติการทดสอบในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ ปรากฏว่า จำนวนข้อสอบ และค่าฟังก์ชั่นสารสนเทศของชุดข้อสอบต่างกัน เนื่องมาจากเมื่อจำนวนข้อสอบ และค่าฟังก์ชั่นสารสนเทศของชุดข้อสอบ เปลี่ยนไป ค่าความถูกต้องในการประมาณค่าความสามารถของผู้สอบ ซึ่งก็รวมไปถึงค่าความคลาดเคลื่อนในการประมาณค่าจะเกิดการเปลี่ยนแปลง กล่าวคือ เมื่อใช้ข้อสอบมากขึ้น ความเชื่อมั่นก็จะมีค่าสูงขึ้น ก็คือ ความคลาดเคลื่อนมาตรฐานก็จะมีค่าลดลง ดังนั้น หากกำหนดค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าต่างกัน จำนวนข้อสอบ และค่าฟังก์ชั่นสารสนเทศของชุดข้อสอบก็จะต่างกันไปด้วย

4. ความสามารถทางคณิตศาสตร์ของผู้สอบในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ จากผลการให้คะแนนที่แตกต่างกัน ซึ่งผู้วิจัยกำหนดไว้ว่า เมื่อผู้สอบมีความสามารถทางคณิตศาสตร์ต่างกัน จะไม่ทำให้ค่าความตรงของการทดสอบต่างกัน ทั้งอาจเป็นเพราะค่าความสามารถของผู้สอบนั้นจะส่งผลต่อความน่าจะเป็นในการตอบข้อสอบได้ถูกต้อง และค่าความน่าจะเป็นดังกล่าวจะส่งผลต่อค่าสารสนเทศของข้อสอบ และค่าสารสนเทศของชุดข้อสอบสูงตามไปด้วย กรณีเช่นนี้จะส่งผลทำให้ค่าความคลาดเคลื่อนมาตรฐานในการวัดความสามารถของผู้สอบมีค่าต่ำลง นั่นคือ มีความถูกต้องในการวัดสูง ในทางกลับกันถ้าความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องมีค่าต่ำก็จะส่งผลให้ค่าสารสนเทศของข้อสอบ และสารสนเทศของชุดข้อสอบมีค่าต่ำ และจะส่งผลให้ค่าความคลาดเคลื่อนมาตรฐานในการวัดมีค่าสูง นั่นคือ มีความถูกต้องในการวัดต่ำนั่นเอง และค่าความถูกต้องในการวัดนั้น เกี่ยวข้องโดยตรงกับค่าความตรงของชุดข้อสอบ ดังนั้น จึงเป็นไปได้ที่เมื่อทดสอบกับผู้สอบที่มีความสามารถสูงจะมีความตรงสูงกว่าเมื่อทดสอบกับผู้ที่มีความสามารถต่ำกว่า อีกประการหนึ่งที่อาจเป็นไปได้ ก็คือ ข้อสอบในคลังข้อสอบนั้น หากพิจารณาโค้งสารสนเทศของชุดข้อสอบแล้ว จะพบว่า เป็นข้อสอบที่เหมาะสมกับผู้สอบที่มีความสามารถปานกลางค่อนข้างสูง ซึ่งหากกำหนดข้อสอบให้มีความเหมาะสมหรือตรงกับความสามารถของผู้สอบแล้ว ก็จะทำให้ค่าของความคลาดเคลื่อนมาตรฐาน ในการวัดต่ำลง (Haladyna & Roid, 1983) เป็นที่น่าสังเกตว่า ความสามารถของผู้สอบนั้นจะไม่สูงมากนัก ทั้งนี้อาจเป็นเพราะเมื่อแบ่งกลุ่มผู้สอบออกไปสามกลุ่ม คือ ความสามารถสูง กลาง และต่ำ ทำให้พิสัยของค่าความสามารถของผู้สอบที่จะนำมาหาค่าสหสัมพันธ์แคบลง พิสัยค่าความสามารถที่แคบลงนี้จะทำให้ความแปรปรวนของค่าความสามารถมีค่าต่ำ ซึ่งการวิจัยครั้งนี้พบว่า ผู้สอบที่มีความสามารถทางคณิตศาสตร์ต่างกัน จะใช้ข้อสอบในการทดสอบไม่เท่ากัน ผู้สอบที่มีความสามารถสูง จะใช้ข้อสอบน้อยกว่าผู้สอบที่มีความสามารถปานกลางและต่ำ กล่าวคือ ผู้สอบที่มีความสามารถสูงก็จะมีความน่าจะเป็นในการตอบข้อสอบได้ถูกต้อง ซึ่งจะส่งผลทำให้ค่าสารสนเทศของข้อสอบและค่าสารสนเทศของชุดข้อสอบสูงตามไปด้วยแต่จะส่งผลทำให้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่ามีค่าต่ำลง และการวิจัยครั้งนี้ ผู้วิจัยได้ใช้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าเป็นเกณฑ์ยุติการทดสอบ เมื่อค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าจากผู้สอบที่มีความสามารถสูงมีค่าต่ำลงเร็วกว่าผู้สอบที่มีความสามารถต่ำ นั่นแสดงว่า ผู้สอบที่มีความสามารถสูงใช้ข้อสอบน้อยข้อกว่าผู้สอบที่มีความสามารถต่ำ