



การศึกษาวิธีการประมาณความสามารถ วิธีการคัดเลือกข้อสอบและเกณฑ์ยุติ การทดสอบ ของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ

A Study of Ability Estimation, Item Selection Methods and Test Termination of Multidimensional Computerized Adaptive Testing

พงษ์พิชิต พรหมสิทธิ์ (Pongpichit Promsit)*

ดร. ไพศาล สุวรรณน้อย (Dr. Paisan Suwannoi)**

ดร. รังสรรค์ มณีเล็ก (Dr. Rangsan Maneelek)***

บทคัดย่อ

การศึกษาค้นคว้าครั้งนี้มีวัตถุประสงค์เพื่อ 1) ศึกษาประสิทธิภาพวิธีการประมาณค่าความสามารถ วิธีการคัดเลือกข้อสอบ และเกณฑ์ยุติการทดสอบ ของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ (MCAT) 2) ศึกษาปฏิสัมพันธ์วิธีการประมาณค่าความสามารถ วิธีการคัดเลือกข้อสอบ และเกณฑ์ยุติการทดสอบ และ 3) เปรียบเทียบประสิทธิภาพวิธีการประมาณค่าความสามารถ 2 วิธี คือ วิธีการประมาณค่าความสามารถแบบความน่าจะเป็นสูงสุด (MLE) และวิธีการประมาณค่าความสามารถแบบเบย์ (MAP) ภายใต้วิธีการคัดเลือกข้อสอบและเกณฑ์ยุติการทดสอบต่างกัน โดยใช้วิธีการคัดเลือกข้อสอบ 2 วิธี คือ วิธีการคัดเลือกข้อสอบแบบความน่าจะเป็นสูงสุด (ML) และวิธีการคัดเลือกข้อสอบแบบเบย์ (BF) และยุติการทดสอบเมื่อความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถน้อยกว่าหรือเท่ากับ 0.3 และน้อยกว่าหรือเท่ากับ 0.4 รวม 8 วิธี (2X2X2) ส่งผลต่อประสิทธิภาพการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติอย่างไร ใช้รูปแบบการวิจัยเชิงทดลอง โดยจำลองข้อมูลผู้สอบ 1,000 คน ด้วยเทคนิคคอมพิวเตอร์แบบมอนติคาร์โล ศึกษาประสิทธิภาพวิธีการประมาณค่าความสามารถ วิธีการคัดเลือกข้อสอบ และเกณฑ์ยุติการทดสอบ โดยวิเคราะห์ความแม่นยำ (Precision) ของค่าประมาณความสามารถผู้สอบ จากรากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) และวิเคราะห์ความถูกต้อง (Accuracy) ของค่าประมาณความสามารถผู้สอบ จากความลำเอียงเฉลี่ย (Bias) วิเคราะห์ปฏิสัมพันธ์วิธีประมาณความสามารถ วิธีคัดเลือกข้อสอบ และเกณฑ์ยุติการทดสอบ โดยพิจารณาความแปรปรวนแบบพหุจำแนก 3 ทาง (Three-way MANOVA) และเปรียบเทียบประสิทธิภาพวิธีประมาณค่าความสามารถ ที่มีวิธีการคัดเลือกข้อสอบและเกณฑ์ยุติการทดสอบต่างกัน โดยพิจารณาความแปรปรวนแบบพหุจำแนกทางเดียว (One-way MANOVA) Multidimensional Computerized Adaptive Testing (MCAT) ผลการวิจัย พบว่า

1. วิธีการประมาณค่าความสามารถแบบเบย์ (MAP) มีประสิทธิภาพสูงกว่าวิธีการประมาณค่าความสามารถแบบความน่าจะเป็นสูงสุด (MLE) วิธีการคัดเลือกข้อสอบแบบความน่าจะเป็นสูงสุด (ML) มีประสิทธิภาพสูงกว่า

คำสำคัญ : ทฤษฎีการตอบสนองของข้อสอบแบบพหุมิติ, การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ

KeyWords : Multidimensional Item Response Theory (MIRT)

* นักศึกษาระดับปริญญาเอก สาขาวิชาการวัดและประเมินผลการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยขอนแก่น

** ผู้ช่วยศาสตราจารย์ประจำสาขาวิชาการวัดและประเมินผลการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยขอนแก่น

*** ผู้อำนวยการสำนักนโยบายและแผนการศึกษาขั้นพื้นฐาน สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน กระทรวงศึกษาธิการ

วิธีการคัดเลือกข้อสอบแบบเบส์ (BF) และเกณฑ์ยุติการทดสอบเมื่อ ความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถน้อยกว่าหรือเท่ากับ 0.3 มีประสิทธิภาพสูงกว่าเกณฑ์ยุติการทดสอบเมื่อความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถน้อยกว่าหรือเท่ากับ 0.4 ซึ่งการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ โดยใช้วิธีการประมาณค่าความสามารถแบบเบส์ (MAP) วิธีการคัดเลือกข้อสอบแบบความน่าจะเป็นสูงสุด (ML) และเกณฑ์ยุติการทดสอบเมื่อ ความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถน้อยกว่าหรือเท่ากับ 0.3 มีประสิทธิภาพการทดสอบสูงสุด

2. วิธีการประมาณความสามารถ วิธีการคัดเลือกข้อสอบ และเกณฑ์ยุติการสอบ ต่างมีปฏิสัมพันธ์ที่ส่งผลต่อความแม่นยำ (Precision) และความถูกต้อง (Accuracy) ของค่าประมาณความสามารถผู้สอบ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

3. วิธีการประมาณค่าความสามารถแบบเบส์ (MAP) มีการประมาณค่าความสามารถผู้สอบได้แม่นยำ (Precision) และถูกต้อง (Accuracy) กว่าวิธีการประมาณค่าความสามารถแบบความน่าจะเป็นสูงสุด (MLE) ตามวิธีการคัดเลือกข้อสอบและเกณฑ์ยุติการทดสอบ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

Abstract

The purposes of this experimental were to 1) study the efficiency of ability estimation methods, item selection methods and test termination according to multidimensional computerized adaptive testing (MCAT) 2) study the interaction of ability estimation methods, item selection methods and test termination, and 3) compare the efficiency of ability estimation with two methods -- maximum likelihood estimate (MLE) and Bayesian estimation method (MAP)--under testing conditions of differences of item selection methods and test termination conditions. There were 8 conditions (2X2X2) consisted of two- ability estimation methods--MLE and MAP methods, two-item selection methods--maximum likelihood (ML) and Bayesian with Fisher's information (BF), two-test termination--standard error of estimation less than or equal to 0.3 and less than or equal to 0.4. This research was based on the data simulation by Markov chain Monte Carlo Methods. There were 1,000 examinees in each condition. Root-mean squared error (RMSE), and BIAS were used as criteria compare the efficiency of methods in this research. In addition, Three-way Multivariate Analysis of Variance (Three-way MANOVA) was conducted to compare means difference of RMSE, and BIAS in order to explore interaction and One-way Multivariate Analysis of Variance (One-way MANOVA) was conducted to compare means difference of RMSE, and BIAS in order to explore efficiency of conditions being studied.

The research findings:

1. The ability estimation methods as Bayesian estimation method was higher efficient than maximum likelihood estimate, the item selection methods as maximum likelihood was higher efficient than the Bayesian with Fisher's information, and the test termination when standard error of estimation less than or equal to 0.3 was higher efficient than the test termination when standard error of estimation less than or equal to 0.4 which the multidimensional computerized adaptive testing by ability estimation methods as Bayesian estimation method, the item selection methods as maximum likelihood and the test termination when standard error of estimation less than or equal to 0.3 with the highest testing efficiency.



2. The interaction of ability estimation methods, item selection methods and test termination, the ability estimation methods and item selection methods, the ability estimation methods and test termination, item selection methods and test termination, there were interactions affecting precision and accuracy of ability estimation of examinees at .01 significant level.

3. The efficiency of ability estimation methods by item selection methods and test termination that Bayesian estimation method was higher efficient than maximum likelihood estimate at .01 significant level.

ความเป็นมาและความสำคัญของปัญหา

วิธีการทดสอบที่นิยมใช้กันโดยทั่วไปในการวัดผลทางการศึกษาคือการทดสอบแบบดั้งเดิม (Conventional Testing) ซึ่งจะใช้แบบทดสอบที่ประกอบไปด้วยกลุ่มข้อสอบที่มีจำนวนมากและจัดเรียงข้อสอบที่แน่นอน ผู้สอบทุกคนจะต้องทำข้อสอบทุกข้อในแบบทดสอบเหมือนกันหมด ซึ่งแบบทดสอบประกอบด้วยข้อสอบที่มีความยากพอเหมาะกับผู้สอบที่มีความสามารถปานกลางเท่านั้น (Vale & Wiess, 1975) ซึ่งการวัดจะมีความถูกต้องแม่นยำที่สุดเมื่อผู้สอบมีความสามารถปานกลางความคลาดเคลื่อนของการวัดจะเพิ่มขึ้นตรงกับความสามารถเบี่ยงเบนจากความสามารถปานกลาง ผู้สอบทุกคนทำข้อสอบทุกข้อที่มีอยู่ในแบบทดสอบ โดยไม่คำนึงว่าข้อสอบเหล่านั้นเหมาะสมกับความสามารถของผู้สอบหรือไม่ เมื่อพิจารณาถึงผลกระทบในเชิงจิตวิทยา ถ้าผู้สอบคนใดได้ตอบข้อสอบที่ยากเกินความสามารถของเขาแล้ว จะทำให้ผู้สอบเกิดความท้อถอยในการทำข้อสอบ และมีแนวโน้มที่จะตอบข้อสอบโดยการเดาสูง ข้อสอบที่ง่ายเกินไปก็จะไม่ท้าทายให้ผู้สอบได้ใช้ความสามารถอย่างเต็มที่ ผู้สอบเกิดความเบื่อหน่าย ผลกระทบเหล่านี้จะส่งผลทำให้ความถูกต้องและแม่นยำของคะแนนที่ได้ลดต่ำลง ทำให้การประมาณความสามารถของผู้สอบคลาดเคลื่อนไป (Weiss, 1974)

จากข้อจำกัดการทดสอบแบบดั้งเดิม นักวัดผลจึงได้คิดค้นวิธีลดความคลาดเคลื่อนในการวัดให้มีความน้อยที่สุด ด้วยวิธีจัดข้อสอบให้เหมาะสมกับระดับความสามารถผู้สอบเรียกว่าการทดสอบแบบปรับเหมาะ (Adaptive Testing) ซึ่งมีลักษณะเด่นคือมุ่งที่จะจัดและคัดเลือกข้อสอบที่มีระดับความยากให้เหมาะสมกับระดับความสามารถของผู้สอบแต่ละคน โดยมีหลักว่าถ้าผู้สอบทำข้อสอบที่

ผ่านมากู้ ข้อสอบข้อถัดไปจะยากขึ้น แต่ถ้าผู้สอบทำข้อสอบที่ผ่านมากู้ ข้อสอบข้อถัดไปจะง่ายลง การทำข้อสอบข้อต่อไปจะปฏิบัติเช่นเดียวกันจนสิ้นสุดการสอบ (Green et al., 1984) ดังนั้น การทดสอบครั้งหนึ่ง ๆ ผู้สอบแต่ละคนไม่จำเป็นต้องทำข้อสอบเหมือนกันทุกข้อ แต่ขึ้นอยู่กับผลการตอบข้อสอบในแบบทดสอบแต่ละคน (Weiss & Kingsbury, 1984)

การทดสอบแบบปรับเหมาะระยะแรกเป็นการสอบที่ใช้กระดาษกับดินสอ (Paper and Pencil Test) โดยให้ผู้สอบเขียนตอบลงในกระดาษคำตอบ ซึ่งการคัดเลือกข้อสอบข้อต่อไปให้ผู้สอบตอบ จะเป็นเรื่องที่ยุ่งยากและไม่สะดวก ต่อมาวิทยาการคอมพิวเตอร์พัฒนาก้าวหน้าขึ้น จึงนำคอมพิวเตอร์มาประยุกต์ใช้กับการทดสอบแบบปรับเหมาะ โดยรวมเอาพัฒนาการของศาสตร์ 2 ด้าน ได้แก่ ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) และเทคโนโลยีซิลิคอนชิป (Silicon Chip Technology) เข้าด้วยกันเพื่อประโยชน์ในการทดสอบเรียกการทดสอบนี้ว่าการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ (Computerized Adaptive Testing: CAT) โดยใช้คอมพิวเตอร์ประมาณค่าความสามารถของผู้สอบ และคัดเลือกข้อสอบให้เหมาะสมกับระดับความสามารถของผู้สอบ ซึ่งสามารถที่จะลดจำนวนข้อสอบลงครึ่งหนึ่งโดยไม่สูญเสียความแม่นยำในการวัด ดังนั้น การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์จึงช่วยลดภาระงานเวลา และค่าใช้จ่ายลงอย่างมาก (Frey & Seitz, 2009)

สำหรับการวิเคราะห์ข้อสอบ เพื่อจัดทำคลังข้อสอบของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ในระยะแรกจะวิเคราะห์ข้อสอบตามทฤษฎีการตอบสนองข้อสอบ โดยใช้โมเดลการตอบสนองข้อสอบแบบเอกมิติ (Unidimensional Item Response Theory Model:

UIRT) โดยมีความเชื่อว่าข้อสอบที่สร้างขึ้นนั้นวัดมิติเชิงเดียว ซึ่งการวัดทางด้านการศึกษาและจิตวิทยาในปัจจุบัน ข้อสอบที่สร้างขึ้นมักวัดมิติความสามารถที่เป็นเชิงซ้อนมากกว่าที่เป็นมิติเชิงเดียว หากวิเคราะห์ข้อสอบตามทฤษฎีการตอบสนองข้อสอบแบบเอกมิติ จะทำให้มีการฝ่าฝืนข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ ที่กำหนดว่าแบบทดสอบที่ใช้วัดความสามารถจะต้องวัดความสามารถด้านเดียวหรือเป็นเอกพันธ์ (Homogeneous Items) หรือมีลักษณะเป็นมิติเดียว (Unidimensional Test) นักวัดผลจึงได้ขยายแนวคิดของทฤษฎีการตอบสนองข้อสอบจากโมเดลการตอบสนองข้อสอบแบบเอกมิติ สู่โมเดลการตอบสนองข้อสอบแบบพหุมิติ (Multidimensional Item Response Theory: MIRT) โดยถือว่าคุณลักษณะแฝงของบุคคลมากกว่า 1 องค์ประกอบ ส่งผลต่อการตอบข้อสอบ พารามิเตอร์ความสามารถผู้สอบ จึงมี 2 พารามิเตอร์ขึ้นไป การพิจารณาความสามารถหลายมิติของบุคคล ช่วยให้โมเดลมีความสอดคล้องกับข้อมูลได้ดียิ่งขึ้น ถ้าผู้สอบใช้ความสามารถหลายมิติที่แตกต่างกันอย่างเป็นระบบเพื่อตอบข้อสอบ ย่อมมีความสัมพันธ์กับความยากและอำนาจจำแนกของข้อสอบรายข้อ และมีผลต่อความน่าจะเป็นในการตอบข้อสอบได้ถูกต้อง

จากแนวคิดดังกล่าว จึงได้มีการขยายรูปแบบการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบเอกมิติ (Unidimensional Computerized Adaptive Testing: UCAT) สู่รูปแบบการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ (Multidimensional Computerized Adaptive Testing: MCAT) ซึ่งเป็นรูปแบบใหม่สำหรับการประเมินความสามารถที่แท้จริงของผู้สอบ เนื่องจากการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ มีอัตราการวัดที่แม่นยำกว่าเมื่อใช้ข้อสอบเท่ากัน สามารถลดจำนวนข้อสอบลงได้มากกว่าการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบเอกมิติ 30-50% และมีประสิทธิภาพสูงกว่าถึง 1.3 เท่า และสามารถลดจำนวนข้อสอบลงได้มากกว่าการทดสอบแบบดั้งเดิมที่กำหนดจำนวนข้อสอบไว้ 70% โดยไม่สูญเสียความแม่นยำ ถ้าใช้วิธีการเลือกข้อสอบแบบสุ่ม (Random) การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ จะมี

ประสิทธิภาพการทดสอบสูงกว่าการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบเอกมิติ 3.7 เท่า (Frey and Seitz, 2009)

จากจุดเด่นการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ ที่สามารถคัดเลือกข้อสอบที่เหมาะสมกับความสามารถของผู้สอบ มีประสิทธิภาพสูง มีความแม่นยำในการประมาณค่าความสามารถผู้สอบ โดยใช้ข้อสอบจำนวนน้อยลงและใช้เวลาทดสอบลดลง ครึ่งหนึ่งของการทดสอบแบบเดิม จึงเป็นเรื่องที่น่าสนใจและนำแนวคิดไปประยุกต์ใช้ดำเนินการทดสอบ เพื่อทราบความสามารถที่แท้จริงของผู้สอบ ช่วยลดความคลาดเคลื่อนจากการทดสอบให้น้อยลงได้ แต่อย่างไรก็ตาม การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ มีวิธีการประมาณค่าความสามารถ วิธีการคัดเลือกข้อสอบ เกณฑ์ยุติการทดสอบที่หลากหลาย ซึ่งการประมาณค่าความสามารถผู้สอบที่มีประสิทธิภาพและนิยมมี 2 วิธี คือ วิธีการประมาณค่าความสามารถแบบความน่าจะเป็นสูงสุด (MLE) และวิธีการประมาณค่าความสามารถแบบเบส์ (MAP) ซึ่งยังไม่มีข้อสรุปเกี่ยวกับปัญหาในการประมาณค่าความสามารถของวิธีทั้งสอง ในทางปฏิบัติ จึงยังใช้วิธีการประมาณค่าความสามารถผู้สอบทั้งแบบความน่าจะเป็นสูงสุดและแบบเบส์ (Frey and Seitz, 2009)

สำหรับวิธีการคัดเลือกข้อสอบ (Item Selection) ของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ เป็นการเลือกข้อสอบจากคลังข้อสอบเพื่อให้ผู้เข้าสอบตอบข้อสอบ หลักการคัดเลือกข้อสอบมีเป้าหมายเพื่อลดความแปรปรวนเวกเตอร์ความสามารถพหุมิติ (Frey and Seitz, 2009) ซึ่งวิธีการคัดเลือกข้อสอบมี 2 วิธี คือ วิธีการคัดเลือกข้อสอบแบบความน่าจะเป็นสูงสุด (ML) และวิธีการคัดเลือกข้อสอบแบบเบส์ (BF) แต่ละวิธีสามารถใช้ร่วมกันกับวิธีการประมาณค่าความสามารถทุกแบบ และสามารถใช่วิธีการคัดเลือกข้อสอบหลายวิธีร่วมกันได้อีกด้วย (Reckase, 2009)

ส่วนเกณฑ์ที่ใช้สำหรับยุติการทดสอบ (Test Termination) ของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ มีหลากหลายลักษณะ เช่น หยุดเมื่อจำนวนข้อสอบที่สอบครบตามจำนวนที่กำหนดไว้ (Fixed length) หยุดเมื่อการวัดความสามารถมีความ



แม่นยำตามเกณฑ์ที่กำหนดไว้ (Variable length) หรือหยุดโดยใช้ทั้งสองเกณฑ์ร่วมกัน ซึ่งการกำหนดจำนวนข้อสอบที่ใช้ในการทดสอบเป็นเกณฑ์ยุติ จะส่งผลต่อความแม่นยำในการประมาณค่าความสามารถที่มีความแปรปรวนระหว่างผู้สอบสูงมาก เนื่องจากผู้สอบบางคนมีความคลาดเคลื่อนมาตรฐานของการวัดจากหนึ่งมิติหรือมากกว่าจะมีขนาดใหญ่ ในขณะที่คนอื่นจะมีขนาดเล็กกว่ามาก และความแม่นยำในการประมาณค่าความสามารถผู้สอบ ยังขึ้นอยู่กับคำตอบข้อสอบและคุณลักษณะคลังข้อสอบ (Frey & Seitz, 2009; Reckase, 2009) ดังนั้น การยุติการทดสอบจากความแม่นยำในการวัดความสามารถ จึงเป็นทางเลือกหนึ่งที่ใช้เป็นเกณฑ์ยุติการทดสอบ โดยจะยุติการทดสอบเมื่อค่าความคลาดเคลื่อนมาตรฐานในการวัดความสามารถตกอยู่ในช่วงความเชื่อถือได้ (Segall, 1996) เช่น การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ของ Newman ใช้เกณฑ์ยุติการทดสอบเมื่อค่าความคลาดเคลื่อนมาตรฐานในการวัดความสามารถมีค่าน้อยกว่าหรือเท่ากับ 0.3, น้อยกว่าหรือเท่ากับ 0.4 และน้อยกว่าหรือเท่ากับ 0.5 (Newman, 1995) และโดยทั่วไปการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์จะยุติการทดสอบเมื่อค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถน้อยกว่าหรือเท่ากับ 0.3 (Thissen, 1990) ซึ่งเกณฑ์ยุติการทดสอบเมื่อค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถน้อยกว่าหรือเท่ากับ 0.3 มีความตรงเชิงสภาพสูงสุด ผู้สอบมีความสามารถสูงจะทำให้ผลการทดสอบมีความตรงเชิงสภาพสูงสุด แต่ใช้จำนวนข้อสอบมากกว่าเกณฑ์ยุติการทดสอบเมื่อค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถน้อยกว่าหรือเท่ากับ 0.4 อยู่ถึงสองเท่า ดังนั้น การทดสอบทั่วไป จึงควรใช้เกณฑ์ยุติการทดสอบด้วยค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถต่ำกว่า 0.4 จะได้ค่าความเชื่อมั่นในการวัดเท่ากับ .9165 แต่ใช้ข้อสอบน้อยลงกว่าครึ่ง (รังสรรค์ มณีเล็ก, 2540) กล่าวได้ว่า การยุติการทดสอบจากความแม่นยำในการวัดโดยใช้เกณฑ์ยุติการทดสอบเมื่อค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถน้อยกว่าหรือเท่ากับ 0.3 และน้อยกว่าหรือ

เท่ากับ 0.4 เป็นที่ยอมรับว่ามีความเหมาะสมในการวัด และมีความสำคัญอย่างมากสำหรับการวิจัยในอนาคต (Reckase, 2009)

จากที่กล่าวมาข้างต้น จะเห็นว่าวิธีการประมาณค่าความสามารถ วิธีการคัดเลือกข้อสอบ และเกณฑ์ยุติการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ สามารถเลือกใช้ได้อย่างหลากหลาย แต่ไม่มีการศึกษาจริงจังว่าวิธีการประมาณค่าความสามารถ วิธีการคัดเลือกข้อสอบ และเกณฑ์ยุติการทดสอบแบบใด มีประสิทธิภาพสูงสุด กล่าวได้ว่า การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ เพื่อให้ได้ประสิทธิภาพสูงสุดลดกระบวนการสอบ ระยะเวลาการสอบ รวมทั้งงบประมาณดำเนินการสอบ จะต้องศึกษาวิธีการประมาณค่าความสามารถผู้สอบ วิธีการคัดเลือกข้อสอบ และเกณฑ์ยุติการทดสอบ เพื่อกำหนดค่าพารามิเตอร์ข้อสอบ ความยาวของแบบทดสอบ ขนาดของคลังข้อสอบ ให้เหมาะสมกับสภาพการสอบ และสามารถนำวิธีการทดสอบไปใช้อย่างมีประสิทธิภาพและเกิดประสิทธิผลสูงสุดต่อไป

วัตถุประสงค์การวิจัย

1. เพื่อศึกษาประสิทธิภาพวิธีการประมาณค่าความสามารถ วิธีการคัดเลือกข้อสอบ และเกณฑ์ยุติการทดสอบ ของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ
2. เพื่อศึกษาปฏิสัมพันธ์วิธีการประมาณค่าความสามารถ วิธีการคัดเลือกข้อสอบ และเกณฑ์ยุติการทดสอบ ของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ
3. เพื่อเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าความสามารถ ที่มีวิธีการคัดเลือกข้อสอบ และเกณฑ์ยุติการทดสอบต่างกัน

วิธีดำเนินการวิจัย

1. การจำลองข้อมูล

จำลองข้อมูล 1) ค่าพารามิเตอร์ความสามารถผู้สอบ 1,000 คน จำนวน 50 ชุด แต่ละชุดมีค่าความสามารถจริง 2 พารามิเตอร์ คือ θ_1 และ θ_2 โดย



มีค่าตั้งแต่ -3.0 ถึง +3.0 2) ค่าพารามิเตอร์ข้อสอบ 300 ข้อ แต่ละข้อจะมีค่าพารามิเตอร์ข้อสอบ ได้แก่ สเกลาร์ ค่าความยากของข้อสอบข้อที่ i (d_i) ครอบคลุมพื้นที่ได้โค้งปกติ มีค่าระหว่าง -3.0 ถึง +3.0 เวกเตอร์ค่าอำนาจจำแนกของข้อสอบ ($\hat{a}_i$) มีค่าระหว่าง 0.2 ถึง 2.5 และค่าความน่าจะเป็นในการตอบข้อสอบข้อที่ i ได้ถูกต้องสำหรับผู้สอบที่มีความสามารถในระดับต่ำสุด (c_i) มีค่าระหว่าง 0.0 ถึง 0.3 และ 3) ผลการตอบข้อสอบของผู้สอบ 1,000 คน จำนวน 50 ชุด โดยมีการให้คะแนนแบบ 2 ค่า คือ ตอบถูกให้ 1 ตอบผิดให้ 0 ซึ่งสารสนเทศที่ได้จากการจำลองข้อมูล ได้ตรวจสอบความถูกต้องข้อมูล ดังนี้

1) ตรวจสอบค่าพารามิเตอร์

การตรวจสอบความถูกต้องข้อมูลจากการจำลอง โดยพิจารณา 1) ค่าพารามิเตอร์ความสามารถผู้สอบ 2) ค่าพารามิเตอร์ข้อสอบ ถ้าค่าพารามิเตอร์ที่ได้ไม่เป็นไปตามที่กำหนด จะดำเนินการจำลองข้อมูลใหม่เพื่อให้ได้ค่าพารามิเตอร์ตามที่กำหนด

2) ตรวจสอบความเป็นพหุมิติ

การตรวจสอบความเป็นพหุมิติของข้อมูล พิจารณาผลการตอบข้อสอบรายข้อจากผู้สอบ 1,000 คน โดยวิเคราะห์องค์ประกอบ (Factor Analysis) แล้วพิจารณาค่าไอเกน (Eigen Value) และน้ำหนักองค์ประกอบ (Factor Loading) ซึ่งเป็นค่าที่ช่วยพิจารณาว่าข้อสอบควรจัดอยู่ในมิติที่ 1 หรือมิติที่ 2 โดยพิจารณาค่าน้ำหนักองค์ประกอบของแต่ละข้อ ถ้าค่าของน้ำหนักองค์ประกอบใดมีค่ามาก (เข้าสู่ +1 หรือ -1) ก็ควรจัดข้อสอบข้อนั้นให้อยู่ในมิติดังกล่าว (พัชร จันทร์เพ็ง, 2551) และกรณีค่าน้ำหนักองค์ประกอบมีค่ากลาง ๆ ทำให้ไม่แน่ใจว่าควรจัดข้อนั้นไว้ในมิติที่ 1 หรือมิติที่ 2 จะทำการจำลองข้อมูลในเงื่อนไขนั้นใหม่ จนกระทั่งได้ค่าน้ำหนักองค์ประกอบเป็นไปตามเงื่อนไขการวิจัย

2. การพัฒนาคัลงข้อสอบ

คัลงข้อสอบ (Item Bank) พัฒนาโดยการจำลองข้อมูลพารามิเตอร์ผู้สอบ พารามิเตอร์ข้อสอบ และผลการตอบแบบให้คะแนน 2 ค่า ดังนี้

1) พารามิเตอร์ความสามารถผู้สอบ 1,000 คน 50 ชุด แต่ละชุดมีค่าความสามารถจริง 2 พารามิเตอร์ คือ θ_1 และ θ_2 มีค่าตั้งแต่ -2.9443 ถึง

+2.9080 สร้างขึ้นโดยสุ่มการแจกแจงโค้งปกติมาตรฐาน มีค่าเฉลี่ยเท่ากับ 0 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1

2) พารามิเตอร์ข้อสอบ 300 ข้อ แต่ละข้อจะมีค่าพารามิเตอร์ข้อสอบ ได้แก่ ค่าความยาก (d_i) ครอบคลุมพื้นที่ได้โค้งปกติ มีค่าอยู่ระหว่าง -2.9915 ถึง +2.9969 ค่าอำนาจจำแนก ($\hat{a}_i$) มีค่าอยู่ระหว่าง 0.2028 ถึง 2.4944 และค่าความน่าจะเป็นในการตอบข้อสอบข้อที่ i ได้ถูกต้องสำหรับผู้สอบที่มีความสามารถในระดับต่ำสุด (c_i) มีค่าอยู่ระหว่าง 0.0001 ถึง 0.2998

3) ผลการตอบข้อสอบของผู้สอบ 1,000 คน 50 ชุด โดยตรวจให้คะแนน 2 ค่า คือ ตอบถูกให้ 1 ตอบผิดให้ 0

3. การพัฒนาโปรแกรมการทดสอบ

ขั้นที่ 1 ศึกษาจุดประสงค์ของการวิจัย วิเคราะห์รูปแบบการทดสอบแบบปรับเหมาะ กำหนดความชัดเจนของการทดสอบ พิจารณาความสัมพันธ์ระหว่างตัวแปรต้น/ตัวแปรตาม กำหนดโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ 2X2X2 วิธี รวม 8 วิธี ดังนี้

วิธีที่ 1 ประเมินค่าความสามารถ MLE คัดเลือกข้อสอบ ML ยุติการทดสอบ $SE \leq 0.3$

วิธีที่ 2 ประเมินค่าความสามารถ MLE คัดเลือกข้อสอบ BF ยุติการทดสอบ $SE \leq 0.3$

วิธีที่ 3 ประเมินค่าความสามารถ MAP คัดเลือกข้อสอบ ML ยุติการทดสอบ $SE \leq 0.3$

วิธีที่ 4 ประเมินค่าความสามารถ MAP คัดเลือกข้อสอบ BF ยุติการทดสอบ $SE \leq 0.3$

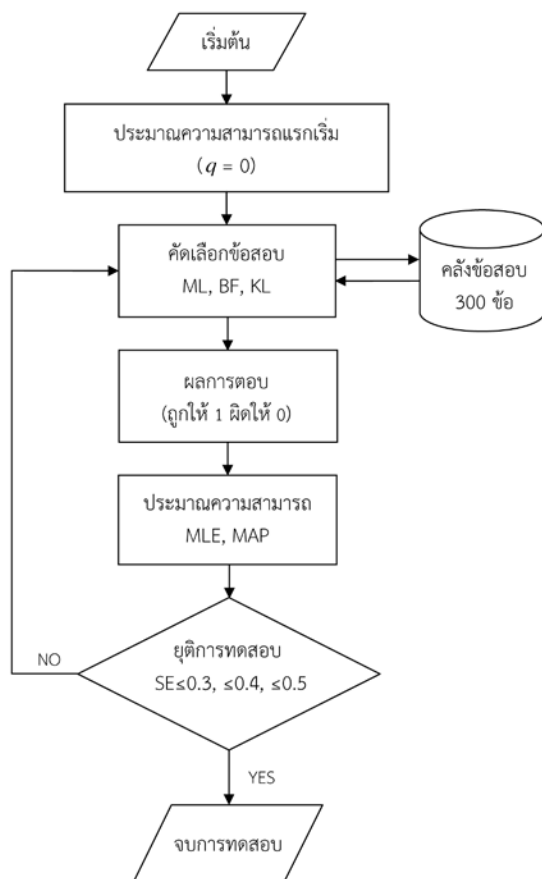
วิธีที่ 5 ประเมินค่าความสามารถ MLE คัดเลือกข้อสอบ ML ยุติการทดสอบ $SE \leq 0.4$

วิธีที่ 6 ประเมินค่าความสามารถ MLE คัดเลือกข้อสอบ BF ยุติการทดสอบ $SE \leq 0.4$

วิธีที่ 7 ประเมินค่าความสามารถ MAP คัดเลือกข้อสอบ ML ยุติการทดสอบ $SE \leq 0.4$

วิธีที่ 8 ประเมินค่าความสามารถ MAP คัดเลือกข้อสอบ BF ยุติการทดสอบ $SE \leq 0.4$

ขั้นที่ 2 เขียนแผนภาพการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ แสดงความสัมพันธ์ของกระบวนการการทดสอบตั้งแต่เริ่มต้นจนถึงสิ้นสุดการทดสอบ ดังนี้



ภาพที่ 1 การทดสอบแบบปรับเหมาะ
ด้วยคอมพิวเตอร์แบบพหุมิติ

ขั้นที่ 3 เขียนโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ ด้วยโปรแกรม MATLAB R2009a โดยเขียนคำสั่งต่างๆ ที่ละขั้นตอนการทำงาน ได้แก่ (1) การประมาณค่าความสามารถผู้สอบ (2) การคัดเลือกข้อสอบให้ผู้สอบตอบ และ (3) การยุติการทดสอบ รวมทั้งหมด 8 โปรแกรม และคำนวณค่า RMSE และค่า Bias เพื่อใช้ในการศึกษาประสิทธิภาพการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ

ขั้นที่ 4 ทดสอบและปรับปรุงโปรแกรมการตรวจสอบขั้นตอนการทำงานโปรแกรมด้วยผู้เชี่ยวชาญติดตามการทำงานที่ละคำสั่งจนกว่าจะจบคำสั่ง ทดสอบโดยการควบคุมเครื่องคอมพิวเตอร์ให้ทำงานตามคำสั่งเมื่อแน่ใจว่าการทำงานโปรแกรมถูกต้องและผลการทดสอบแม่นยำเพียงพอสำหรับเก็บข้อมูลตัวแปรตามได้ครบถ้วนจึงนำไปใช้ทดสอบในงานวิจัยต่อไป

เครื่องมือที่ใช้ในการวิจัย

โปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ ภายใต้วิธีการประมาณค่าความสามารถ 2 วิธี คือ วิธี MLE และวิธี MAP วิธีการคัดเลือกข้อสอบ 2 วิธี คือ วิธี ML และวิธี BF และยุติการทดสอบเมื่อค่า $SE \leq 0.3$ และ $SE \leq 0.4$ รวม 8 วิธี (2X2X2)

การวิเคราะห์ข้อมูล

1. วิเคราะห์ความแม่นยำ (Precision) ค่าประมาณความสามารถ จาการากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Square Error: RMSE) ความถูกต้อง (Accuracy) ค่าประมาณความสามารถผู้สอบ จากความลำเอียงเฉลี่ย (Average Bias: Bias)
2. วิเคราะห์ปฏิสัมพันธ์วิธีการประมาณความสามารถ วิธีการคัดเลือกข้อสอบ และเกณฑ์ยุติการทดสอบ จากความแปรปรวนพหุจำแนก 3 ทาง (Three-way MANOVA)
3. เปรียบเทียบประสิทธิภาพวิธีการประมาณค่าความสามารถ ที่มีวิธีการคัดเลือกข้อสอบและเกณฑ์ยุติการทดสอบต่างกัน จากความแปรปรวนพหุจำแนกทางเดียว (One-way MANOVA)

ผลการวิจัยและอภิปรายผลการวิจัย

1. วิธีประมาณค่าความสามารถแบบ MAP มีความแม่นยำและถูกต้องสูงกว่าวิธีประมาณค่าความสามารถแบบ MLE วิธีคัดเลือกข้อสอบแบบ ML มีความแม่นยำและถูกต้องสูงกว่าวิธีคัดเลือกข้อสอบแบบ BF เกณฑ์ยุติการทดสอบเมื่อความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถน้อยกว่าหรือเท่ากับ 0.3 มีความแม่นยำและถูกต้องสูงกว่าเกณฑ์ยุติการทดสอบเมื่อความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถน้อยกว่าหรือเท่ากับ 0.4 ซึ่งการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ โดยใช้วิธีประมาณค่าความสามารถแบบ MAP วิธีคัดเลือกข้อสอบแบบ ML และยุติการทดสอบเมื่อความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถน้อยกว่าหรือเท่ากับ 0.3 มีประสิทธิภาพการทดสอบสูงสุด นั่นคือ



การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ จะให้ประสิทธิภาพสูงสุด เมื่อใช้วิธีการประมาณค่าความสามารถแบบ MAP และใช้วิธีการคัดเลือกข้อสอบแบบ ML และยุติการทดสอบเมื่อความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถน้อยกว่าหรือเท่ากับ 0.3

2. มีปฏิสัมพันธ์ระหว่างวิธีประมาณค่าความสามารถ วิธีคัดเลือกข้อสอบ เกณฑ์ยุติการทดสอบ การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 นั่นคือ วิธีประมาณค่าความสามารถ วิธีคัดเลือกข้อสอบ และเกณฑ์ยุติการทดสอบ มีปฏิสัมพันธ์ที่ส่งผลต่อความแม่นยำและความถูกต้องของค่าประมาณความสามารถผู้สอบ แสดงดังตารางที่ 1

ตารางที่ 1 ปฏิสัมพันธ์วิธีประมาณค่าความสามารถ วิธีคัดเลือกข้อสอบ และเกณฑ์ยุติการทดสอบ

แหล่งความแปรปรวน	ตัวแปรตาม	SS	df	MS	F	p
วิธีประมาณค่า	RMSE1	77.672	4	19.418	141.531	.000
ความสามารถ X	RMSE2	257.559	4	64.390	389.194	.000
วิธีคัดเลือกข้อสอบ X เกณฑ์	BIAS1	12.709	4	3.177	6.504	.000
ยุติการสอบ	BIAS2	42.656	4	10.664	18.390	.000

3. วิธีประมาณค่าความสามารถแบบ MAP สามารถประมาณค่าความสามารถผู้สอบได้แม่นยำและถูกต้องกว่าวิธีประมาณค่าความสามารถแบบ MLE อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 กล่าวได้ว่า การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ จะมี

ประสิทธิภาพในการประมาณค่าความสามารถได้แม่นยำและถูกต้องมากที่สุด เมื่อใช้วิธีการประมาณค่าความสามารถแบบ MAP ไม่ว่าจะใช้วิธีคัดเลือกข้อสอบและเกณฑ์ยุติการทดสอบแบบใด แสดงดังตารางที่ 2-3

ตารางที่ 2 เปรียบเทียบประสิทธิภาพวิธีประมาณค่าความสามารถ ตามวิธีการคัดเลือกข้อสอบ

การจำลองข้อมูล	วิธี	ตัวแปรตาม	Mean	S.D.	F	P
วิธีการคัดเลือกข้อสอบ	ML	MLE	1.112	0.301	7176.5	.000
		MAP	0.505	0.252		
		MLE	1.594	0.568	7428.666	.000
		MAP	0.672	0.141		
		MLE	0.057	0.403	283.2843	.000
		MAP	0.226	0.372		
	BF	MLE	-0.072	0.683	1815.127	.000
		MAP	-0.619	0.162		
		MLE	1.704	0.531	8294.625	.000
		MAP	0.610	0.389		
		MLE	2.268	0.788	10274.72	.000
		MAP	0.737	0.251		
		MLE	0.626	0.849	594.7674	.000
		MAP	0.151	0.647		
		MLE	-0.488	1.334	70.19558	.000
		MAP	-0.697	0.276		



ตารางที่ 3 เปรียบเทียบประสิทธิภาพวิธีประมาณค่าความสามารถ ตามเกณฑ์ยุติการทดสอบ

การจำลองข้อมูล	วิธี	ตัวแปรตาม	Mean	S.D.	F	P
เกณฑ์ยุติการทดสอบ	MLE	RMSE ₁	1.551	0.586	6558.994	.000
	MAP		0.572	0.310		
	MLE	RMSE ₂	1.767	0.620	8976.389	.000
	MAP		0.667	0.141		
	MLE	Bias ₁	0.369	0.801	44.73052	.000
	MAP		0.255	0.475		
	MLE	Bias ₂	-0.274	0.898	375.1011	.000
	MAP		-0.597	0.159		
	MLE	RMSE ₁	1.635	0.581	7202.663	.000
	MAP		0.593	0.339		
	MLE	RMSE ₂	1.963	0.574	12527.76	.000
	MAP		0.729	0.189		
	MLE	Bias ₁	0.416	0.855	140.9135	.000
	MAP		0.196	0.546		
	MLE	Bias ₂	-0.434	0.966	163.0779	.000
	MAP		-0.665	0.215		

ข้อเสนอแนะที่ได้จากการวิจัย

1. วิธีการประมาณค่าความสามารถแบบ MAP มีการประมาณค่าความสามารถผู้สอบได้แม่นยำและถูกต้องกว่าวิธีการประมาณค่าความสามารถแบบ MLE แต่วิธีการ MAP จำเป็นต้องทราบความสามารถที่แท้จริงหรือสารสนเทศเดิมผู้สอบ (Prior Information) เพื่อประมาณค่าความสามารถใหม่ ซึ่งสถานการณ์การทดสอบจริงจะไม่สามารถทราบความสามารถที่แท้จริงผู้สอบได้ ในทางปฏิบัติจึงมักใช้สารสนเทศเดิมของผู้สอบ เช่น คะแนนสอบวัดผลสัมฤทธิ์ทางการเรียน คะแนน GPA มาใช้ในการกำหนดความสามารถที่แท้จริงผู้สอบ

2. การวิจัยครั้งนี้เป็นการจำลองข้อมูล ซึ่งค่าพารามิเตอร์ผู้สอบ ค่าพารามิเตอร์ข้อสอบ และผลการตอบข้อสอบ ถูกจัดไว้ให้ครอบคลุมและเป็นไปตามเงื่อนไขที่กำหนด แต่การนำไปใช้จริงอาจจะไม่เป็นไปตามข้อค้นพบตามผลการวิจัยนี้ จึงควรมีการศึกษาประสิทธิภาพ

การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติในบริบทการทดสอบจริง โดยหาค่าพารามิเตอร์ผู้สอบ ค่าพารามิเตอร์ข้อสอบ ผลการตอบข้อสอบ และพัฒนาคลังข้อสอบจากการทดลองใช้จริง โดยนำรูปแบบการทดสอบแบบปรับเหมาะแต่ละรูปแบบไปใช้จริง เพื่อยืนยันข้อค้นพบที่ได้จากการวิจัยครั้งนี้

3. การวัดทางการศึกษาและจิตวิทยาในปัจจุบัน ลักษณะโครงสร้างของแบบทดสอบมีลักษณะเป็นองค์ประกอบเชิงซ้อนภายใต้การวัดหลายมิติ และรูปแบบแบบทดสอบจะเป็นรูปแบบผสม ประกอบด้วยข้อสอบเลือกตอบหลายตัวเลือก ตรวจให้คะแนน 2 ค่า และข้อสอบอัตนัย ตรวจให้คะแนนหลายค่า ในฉบับเดียวกัน ซึ่งการศึกษาประสิทธิภาพการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติในอนาคต จำเป็นจะต้องคำนึงถึงลักษณะมิติความสามารถและลักษณะของแบบทดสอบดังกล่าวด้วย

เอกสารอ้างอิง

- พัชรี จันทรเพ็ง. (2551). การเปรียบเทียบคุณภาพของวิธีการเชื่อมโยงคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ ภายใต้การหมุนแกน โครงสร้างเชิงมิติและระดับความสัมพันธ์ที่แตกต่างกัน. วิทยานิพนธ์ปริญญาครุศาสตรดุษฎีบัณฑิต สาขาการวัดและประเมินผลการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.
- รังสรรค์ มณีเล็ก. (2540). ผลของตัวแปรบางตัวต่อความเที่ยงตรงเชิงสภาพและจำนวนข้อสอบที่ใช้ในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์. วิทยานิพนธ์ปริญญาการศึกษาดุษฎีบัณฑิต สาขาการวัดและประเมินผลการศึกษา บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒประสานมิตร.
- Frey, A., & Seitz, N. N. (2009). Multidimensional adaptive testing in educational and Psychological measurement: Current state and future challenges. **Studies in Educational Evaluation**, 35, 89-94.
- Green, B. F. & Other. (1984). Technical guidelines for assessing computerized adaptive tests. **Journal of Educational Measurement**, 21, 347 - 360.
- Newman, L. S. (1995). Content validity of a computerized adaptive licensing and certification examination: A comparison of content-balancing methods. **Temple University. ProQuest Dissertations and Theses**. Retrived January 10, 2010, from <http://search.proquest.com/clocview/304232692?accountid=27797>.
- Reckase, Mark D. (2009). **Multidimensional Item Response Theory**. New York: Springer.
- Segall D. O. (1996). Multidimensional adaptive testing. **Psychometrika**, 61(2), 331-354.
- Thissen, David. (1990). **Reliability and Measurement Precision**. in **Computerized Adaptive Testing: A Primer**. Wairer, H. (Ed.) (pp. 161-186). New Jersey: Lawrence Erlbaum.
- Vale, C. D. & Weiss, D. J. (1975). **A Simulation study of strapdative ability testing**. Research Report. Ma: Department of Psychology, University of Minnesota.
- van der Linden, W. J. & Pashley, P. J. (2010). Item selection and ability estimation in adaptive Testing. In Wim J. van der Linden & Cees AW Glas (Eds.). **Elements of adaptive testing**. New York: Springer.
- Weiss, D. J. (1974). Strategies of adaptive ability measurement. **Research Report**. (pp. 74-75). Ma: Department of Psychology, University of Minnesota.
- Weiss, D. J. & Kingsbury, G. G. (1984). Application of computerized adaptive testing of educational problems. **Journal of Educational Measurement**, 21, 361-375.