การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ
(Multidimensional Computerized Adaptive Testing : MCAT)

การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ (Multidimensional Computerized Adaptive Testing : MCAT) เป็นการผสมผสานแนวคิดระหว่างทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (Multidimensional Item Response Theory : MIRT) กับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ (Computerized Adaptive Testing : CAT) ซึ่งเป็นการทดสอบที่มีประสิทธิภาพสูง มีความแม่นยำ และลดจำนวนข้อสอบ

1. ความเป็นมาและหลักการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ(Multidimensional Computerized Adaptive Testing)

เป็นเวลามากกว่า 10 ปี ที่การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ (Computerized Adaptive Testing) หรือที่เรียกว่า (CAT) ได้รับความนิยม และกล่าวถึงในความสำเร็จของวิธีการทดสอบนี้ ซึ่งการทดสอบแบบปรับเหมาะนั้นมีความสามารถมากกว่าการทดสอบแบบดั้งเดิมที่เขียนลงบนกระดาษ คำตอบ (paper-and-pencil tests) ซึ่งเป็นการเพิ่มความแม่นยำในการวัด ลดระยะเวลาในการทดสอบ วิธีใช้ได้มาตรฐานและมีความยืดหยุ่นในการจัดตารางการสอบของผู้สอบ การทดสอบแบบปรับเหมาะส่วนใหญ่จะใช้การเลือกข้อคำถามและการคำนวณการให้คะแนนโดยใช้พื้นฐานของทฤษฎีการตอบสนองข้อสอบ (item response theory : IRT) ซึ่งในขณะนี้เทคนิคดังกล่าวใช้กับการทดสอบแบบปรับเหมาะภายใต้ข้อตกลงเบื้องต้นที่การทดสอบจะต้องเป็นรูปแบบเอกมิติ (Unidimensionality) (Segall. 1996 : 331-354) และเนื่องจากการตอบสนองข้อสอบแบบเอกมิติ ในบางครั้ง จะไม่เหมาะสมกับการทดสอบจริง ดังนั้นจึงมีการพัฒนาแนวคิดเกี่ยวกับทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (Multidimensional Item Response Theory : MIRT) ซึ่งมีแนวคิดมาจาก 2 กลุ่ม คือ Spearman (1927) and Thurstone (1947) และ Lazarsfeld (1950) (พัชรี จันทร์เพ็ง. 2550 : 22) ซึ่งในขณะที่ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติเริ่มได้รับการยอมรับและการทดสอบแบบปรับเหมาะกำลังได้รับความนิยมอย่างมากในการนำไปปฏิบัติจึงเกิดการรวมกันของทั้งสองวิธีนี้เรียกว่า การทดสอบปรับเหมาะแบบพหุมิติ (Multidimensional Adaptive Testing : MAT) (Segall. 2010 : 57-75) และเมื่อนำคอมพิวเตอร์มาใช้ในกระบวนการทดสอบจึงเรียกการทดสอบนี้ว่า การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ (Multidimensional Computerized Adaptive Testing : MCAT) ซึ่งวิธีการทดสอบดังกล่าวนี้เป็นการทดสอบที่มีประสิทธิภาพในการวัดสูง ไม่ว่าจะเป็นความแม่นยำในการวัดการลดความยาวของข้อสอบ และเป็นการวัดที่ตรงกับสภาพจริงของลักษณะข้อสอบ

การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ได้รับความนิยมมากทั้งในการทดสอบทางการศึกษาและจิตวิทยา และเป็นวิธีการเฉพาะที่ใช้ในการประเมินความสามารถและคุณลักษณะแฝงอื่นๆ ในการเลือกข้อสอบจากคลังข้อสอบที่ให้ผู้สอบได้รับข้อสอบตามผลการตอบข้อสอบของข้อก่อนหน้านี้จุดมุ่งหมายของขั้นตอนการเลือกข้อสอบ คือ การเลือกข้อสอบให้เหมาะสมกับระดับลักษณะของผู้เข้าสอบ ซึ่งในปัจจุบันการทดสอบจำนวนมากที่มีลักษณะเหมือนกัน เช่น การสอบภาษาอังกฤษในลักษณะเป็นภาษาต่างประเทศ (TOEFL) ข้อสอบวัดเชาว์ปัญญาทั่วไปก่อนเข้าศึกษาในระดับบัณฑิตศึกษา (the Graduate Record Examinations : GREs) หรือข้อสอบที่ใช้วัดความสามารถของผู้ที่ต้องการเข้าศึกษาในระดับปริญญาโทและปริญญาเอกสาขาบริหารธุรกิจ (Graduate Management Admission Test: GMAT) ที่ใช้การบริหารการสอบแบบปรับเหมาะ

ด้วยคอมพิวเตอร์ นอกจากนี้ ในปี 2012 องค์กรความร่วมมือและพัฒนาทางเศรษฐกิจ (OECD) มีแผนที่จะนำเอาการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์มาใช้ในการประเมินสมรรถนะของนักเรียนในโครงการประเมินนักเรียนในระดับนานาชาติ (Programme for International Student Assessment : PISA) (Frey and Seitz. 2009 :89-90)

แนวคิดของการทดสอบแบบปรับเหมาะแบบพหุมิติเป็นที่รู้จักกันมาในช่วง 20 ปีที่ผ่านมา ในปี 1987 Bloxom และ Vale ได้นำเสนอวิธีการที่ปรับปรุงใหม่เกี่ยวกับการทดสอบแบบปรับเหมาะแบบพหุมิติ โดยใช้การคัดเลือกข้อสอบแบบเบส์ปรับปรุงใหม่ของ Owen ที่นำเสนอไว้ในปี 1975หลังจากนั้นในปี 1992 Tam ได้ศึกษาเปรียบเทียบวิธีการคัดเลือกข้อสอบในการทดสอบแบบปรับเหมาะแบบพหุมิติระหว่างวิธีแบบเบส์ปรับปรุงใหม่ของ Owen กับวิธีการอื่น ๆ ในประเด็นความแม่นยำในการวัดสารสนเทศของข้อสอบและเวลาในการคำนวณ ต่อมาในปี 1996 Segall ได้นำเสนอวิธีการแบบเบส์สำหรับการทดสอบแบบปรับเหมาะแบบพหุมิติ ซึ่งในเวลาต่อมาวิธีการแบบเบส์เป็นวิธีการที่ได้รับความสนใจมาก วิธีการนี้ใช้ความรู้เกี่ยวกับโครงสร้างของความแปรปรวนร่วมระหว่างมิติของการวัดสำหรับการเลือกข้อสอบและการประมาณค่าความสามารถ ส่งผลให้จำนวนข้อสอบที่ใช้ลดลง ซึ่งเป็นคุณสมบัติที่จำเป็นของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ แนวคิดของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ เป็นแนวทางใหม่ในการประเมินความสามารถด้วยเหตุผล 2 ประการ ประการแรก เมื่อเปรียบเทียบกับการใช้โมเดลการวัดแบบเอกมิติ การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติย่อมมีข้อจำกัดที่เหมาะสมระหว่างโมเดลเชิงทฤษฎีที่ประกอบกันขึ้นมาจากหลากหลายแนวคิดและโมเดลการวัดในเชิงสถิติด้วยการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติกับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบเอกมิติ และการทดสอบแบบดั้งเดิม (Frey and Seitz.2009 : 89-90)

นอกจากนี้ Frey และ Cartensen (2009 : 59) ได้กล่าวถึง การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติว่าเป็นการนำเอาหลักการของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์มาขยายขอบเขตการวัดของข้อสอบจากการวัดแบบเอกมิติ เป็นข้อสอบแต่ละข้อวัดพร้อมกันในหลาย ๆ มิติ และถ้าใช้จำนวนข้อสอบที่เท่ากัน การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติจะเป็นการทดสอบที่มีประสิทธิภาพสูงกว่าการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบเอกมิติ และการทดสอบแบบดั้งเดิม เนื่องจากการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติมีอัตราการวัดที่แม่นยำกว่าเมื่อใช้ข้อสอบเท่ากัน ดังนั้น การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติจึงสามารถที่จะลดจำนวนข้อสอบลงได้มากกว่าการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบเอกมิติประมาณ 30–50% และลดจำนวนข้อสอบได้มากกว่าการทดสอบแบบดั้งเดิมที่กำหนดจำนวนข้อสอบไว้ประมาณ 70% โดยไม่สูญเสียความแม่นยำ และ Frey และ Seitz ได้กล่าวถึงการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติว่ามีประสิทธิภาพสูงกว่าการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบเอกมิติ 1.3 เท่า อย่างไรก็ตาม Frey และ Seitz ได้ระบุว่าถ้าใช้การเลือกข้อสอบแบบสุ่มการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติจะมีประสิทธิภาพสูงกว่าการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบเอกมิติ 3.7 เท่า (Frey and Seitz. 2009 : 81-93)

โดยสรุป การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์เป็นการทดสอบที่ผู้สอบแต่ละคนได้ข้อสอบแตกต่างชุดกันด้วยเกณฑ์การคัดเลือกข้อสอบ ที่มีความเหมาะสมกับระดับความสามารถของผู้สอบ ทำให้ผู้สอบจะไม่มีความรู้สึกว่าข้อสอบมีความยากมากเกินความสามารถของผู้สอบด้วยการออกแบบคลังข้อสอบที่มี

ข้อสอบจำนวนมากและมีเนื้อหาครอบคลุมสิ่งที่ต้องการวัด นอกจากนี้ การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ยังสามารถตรวจและรายงานผลได้อย่างรวดเร็ว ทำให้ตอบสนองความต้องการของผู้สอบที่ต้องการทราบผลทันทีหลังการทดสอบ การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ จึงเป็นการทดสอบที่เหมาะที่จะนำไปใช้ในการจัดการเรียนการสอนเพื่อตอบสนองผู้เรียนอย่างแท้จริง

2. หลักการของการทดสอบแบบปรับเหมาะ (Principle of Adaptive Testing)

การทดสอบแบบปรับเหมาะ (Adaptive Testing or Tailored Testing) เป็นการทดสอบที่มุ่งวัดคุณลักษณะที่ต้องการ โดยการคัดเลือกข้อสอบตามความสามารถของผู้สอบในระหว่างกระบวนการทดสอบ การคัดเลือกข้อสอบจะพิจารณาจากผลการตอบข้อสอบหรือข้อคำถามที่มีมาก่อนถ้าตอบข้อสอบข้อนั้นได้ถูกต้อง ข้อสอบข้อต่อไปก็จะยากขึ้น แต่ถ้าตอบข้อสอบข้อนั้นผิด ข้อสอบข้อต่อไปก็จะง่ายลง (Weiss. 1983 : 1-7 ; Weiss and Schleisman. 1999 : 138-150) แบบทดสอบที่ใช้ในการทดสอบแบบปรับเหมาะจึงเป็นแบบทดสอบที่ประกอบไปด้วยชุดของข้อสอบที่แตกต่างกันสำหรับการสอบของแต่ละบุคคล ซึ่งเป็นลักษณะที่ตรงกันข้ามกับแบบทดสอบแบบประเพณีนิยม (Conventional Tests) ที่ผู้สอบทุกคนจะถูกสอบด้วยแบบทดสอบชุดเดียวกัน ซึ่งแบบทดสอบประเพณีนิยมอาศัยหลักการตามทฤษฎีการทดสอบแบบมาตรฐานเดิม (Classical Test Theory : CTT) คุณภาพของข้อสอบได้มาก็ต่อเมื่อผู้สอบทุกคนทำข้อสอบทุกข้อในแบบทดสอบชุดเดียวกัน ผลการทดสอบจะได้ค่าความสามารถของผู้สอบกระจายเป็นช่วงกว้างบนเส้นต่อเนื่องของค่าความสามารถตามจุดประสงค์การวัดของแต่ละแบบทดสอบนั้น

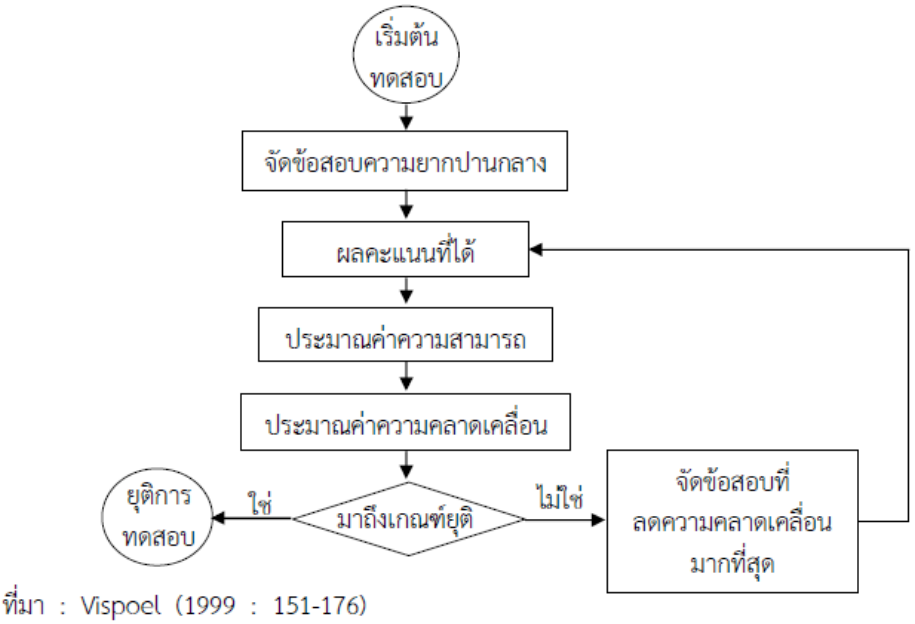
ข้อสอบในแบบทดสอบประเพณีนิยมมักกำหนดค่าความยากปานกลางประมาณ 0.5 ผู้สอบระดับความสามารถบริเวณช่วงค่าความยาก 0.5 จะได้รับการวัดผลอย่างแม่นยำ แบบทดสอบประเพณีนิยม จึงใช้ได้ดีกับผู้สอบระดับความสามารถปานกลางเท่านั้น อีกประการหนึ่ง การให้ผู้สอบทุกคนต้องทำข้อสอบจำนวนเท่า ๆ กัน ทำให้มีข้อสอบจำนวนน้อยที่เหมาะสมกับผู้สอบระดับความสามารถสูงและผู้สอบระดับความสามารถต่ำ ความแม่นยำในการวัดสำหรับผู้สอบเหล่านี้จึงลดลงอย่างรวดเร็วเมื่อข้อสอบมีความยากเกินไปสำหรับผู้สอบความสามารถต่ำหรือง่ายเกินไปสำหรับผู้สอบความสามารถสูงสารสนเทศของผู้สอบจึงน้อยลงตามไปด้วย เพราะฉะนั้น ข้อสอบยากเกินไปหรือง่ายเกินไปเมื่อจัดให้ผู้สอบจะไม่วัดความสามารถของผู้สอบได้ถูกต้อง (Weiss. 1974 ; 1985 : 774-789) มีผลต่อความเชื่อมั่นในภาพรวมและให้ความเที่ยงตรงลดลง

การจำกัดเวลาในการทดสอบเพื่อความสะดวกของผู้คุมสอบ อาจนำมาซึ่งความคลาดเคลื่อนของคะแนนแบบทดสอบได้ เนื่องจากผู้สอบบางคนตอบข้อสอบได้ในเวลาจำกัดในขณะที่บางคนทำไม่ได้ ผลคือการจำกัดเวลาจะส่งผลกระทบต่อผู้สอบแตกต่างกัน การประมาณขนาดความคลาดเคลื่อนของคะแนนแบบทดสอบทำได้ไม่แน่นอน (Weiss. 1974)

จากเหตุผลดังกล่าว มีการศึกษาวิธีการทดสอบอื่น ๆ นอกเหนือจากการทดสอบประเพณีนิยม เพื่อเพิ่มความเชื่อมั่นและความเที่ยงตรงของคะแนนการทดสอบ วิธีที่ได้รับความสนใจคือ การทดสอบแบบปรับเหมาะตามหลักการตามทฤษฎีการตอบข้อสอบ (Item Response Theory : IRT) จากแนวคิดว่า ข้อสอบที่จัดให้แก่ผู้สอบแต่ละคนมีค่าความยากเหมาะสมกับความสามารถของผู้สอบคนนั้นทำให้ความน่าจะเป็นของการตอบถูกต้องหรือตอบผิดมีค่าเท่ากับ 0.5 ดังนั้น ข้อสอบที่จัดให้แก่ผู้สอบจะวัดระดับความสามารถของผู้สอบคนนั้นอย่างแม่นยำ มีความเฉพาะเจาะจงมากกว่าการทดสอบประเพณีนิยมซึ่งให้เพียงระดับความสามารถโดยเฉลี่ยของกลุ่มผู้สอบ

ภายใต้กลยุทธ์ของการทดสอบแบบปรับเหมาะ การใช้ข้อสอบจำนวนมากหรือน้อยนั้นขึ้นอยู่กับระดับความสามารถของผู้สอบ ข้อสอบความยากเหมาะสมกับความสามารถของผู้สอบจะลดผลที่เป็นอุปสรรคและการเดา รวมทั้งจำนวนข้อสอบต่ำกว่าหรือสูงกว่าระดับความสามารถของผู้สอบจะใช้น้อยลง ลดความเบื่อหน่ายของผู้สอบในการทำแบบทดสอบ โดยทั่วไป การทดสอบแบบปรับเหมาะจะไม่จำกัดเวลา อนุญาตให้ผู้สอบได้ใช้เวลาตามต้องการ การจัดจำนวนข้อสอบให้มีความเฉพาะในแต่ละบุคคล ทำให้ผู้สอบบางคนลดเวลาในการทำแบบทดสอบ นอกจากนี้ เวลาในการตอบข้อสอบสามารถใช้เป็นข้อมูลในการวัดทางจิตวิทยาได้ Weiss and Betz (Weiss. 1974 ; citing Weiss and Betz.1973) กล่าวว่า การทดสอบแบบปรับเหมาะสามารถลดเวลาโดยไม่ลดความเชื่อมั่นและความเที่ยงตรงและอาจจะเพิ่มความเชื่อมั่นและความเที่ยงตรงได้

การดำเนินการทดสอบแบบปรับเหมาะจะเกี่ยวข้องกับคำถามสำคัญ 3 คำถาม ได้แก่ จะเลือกข้อสอบที่ใช้ในการเริ่มต้นการสอบได้อย่างไร (how to start) จะเลือกข้อสอบข้อถัดไปได้อย่างไร (how to continue) และรู้ได้อย่างไรว่าจะยุติการสอบ (how to stop) (Thissen and Mislevy.1990 ; Wainer. 1990 : 1-21) คำถามดังกล่าวจะไม่พบในการทดสอบแบบประเพณีนิยมเนื่องจากในการเริ่มต้นทำแบบทดสอบมักจะเริ่มจากการทำข้อสอบข้อที่ 1 ข้อต่อ ๆ ไป จนถึงข้อสุดท้ายแต่ในการทดสอบแบบปรับเหมาะ การเลือกข้อสอบที่ใช้ในการประมาณค่าความสามารถเบื้องต้นหรือเป็นขั้นตอนของการเริ่มการทดสอบเป็นที่สำคัญและมีผลต่อการประมาณค่าความสามารถของผู้สอบเพื่อการเลือกข้อสอบข้อต่อไป รวมถึงการยุติการสอบ และการประมาณค่าความสามารถสุดท้ายของผู้สอบการเริ่มต้นการสอบ การดำเนินการสอบและการยุติการสอบจึงเป็นสิ่งที่ต้องพิจารณาเพื่อให้การทดสอบแบบปรับเหมาะมีประสิทธิภาพสูงสุดและมักจะเป็นประเด็นที่ใช้ในการศึกษาวิจัยพัฒนาเกี่ยวกับการทดสอบแบบปรับเหมาะ ซึ่งในการทดสอบแบบปรับเหมาะนั้น อาศัยแนวคิดตามทฤษฎีการตอบข้อสอบ (Item Response Theory : IRT) โดยทั่วไปนิยมใช้รูปแบบ 3 พารามิเตอร์ ข้อสอบต้องทราบค่าพารามิเตอร์และรวบรวมข้อสอบทั้งหมดไว้ในคลังข้อสอบ ขั้นตอนการทดสอบแบบปรับเหมาะสรุป ดังนี้



ที่มา : Vispoel (1999 : 151-176)

ภาพประกอบ ขั้นตอนการทดสอบแบบปรับเหมาะ

ภาพประกอบ สรุปขั้นตอนการทดสอบแบบปรับเหมาะ เริ่มต้นการทดสอบทำโดยเลือกข้อสอบจากคลังข้อสอบกำหนดความยากของข้อสอบที่ระดับปานกลางหรือความยากของข้อสอบใกล้เคียงกับความสามารถของผู้สอบตามสารสนเทศเดิม ผลคะแนนที่ได้นำไปคำนวณหาค่าประมาณความสามารถของผู้สอบ เพื่อใช้ค่าประมาณความสามารถนี้ทำการเลือกข้อสอบข้อที่ให้ความคลาดเคลื่อนน้อยที่สุดและจัดให้ผู้สอบ นั่นหมายถึง ถ้าตอบข้อสอบถูกต้อง ข้อสอบข้อต่อไปมีระดับความยากเพิ่มขึ้น ถ้าตอบข้อสอบผิด ข้อสอบข้อต่อไปมีระดับความยากลดลง การทดสอบดำเนินต่อเนื่องไปเรื่อยๆ และสิ้นสุดเมื่อความคลาดเคลื่อนของการประมาณค่าความสามารถไม่เกินเกณฑ์กำหนด หรือค่าประมาณความสามารถอยู่ในระดับยอมรับได้ หรือเมื่อใช้ข้อสอบตามจำนวนต้องการหรือ ดังนั้น ข้อสอบที่จัดให้แก่ผู้สอบกล่าวได้ว่าไม่ง่ายเกินไปหรือยากเกินไปกว่าความสามารถของผู้สอบ เหมาะกับการวัดระดับความสามารถของผู้สอบแต่ละคน

ตาราง สรุปประเด็นในงานวิจัยที่เกี่ยวข้องกับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์

ผู้วิจัย	ประเด็นที่ศึกษา				
	Item Bank	Item Selection	Ability Estimate	Termination Criteria	Repeated answer
โสพส (2545)	$0 < a \leq +3$ $-3 < b < +3$ $c < 0.3$	b กับ $\hat{\theta}$	Bayesian	SEE < 0.45 หรือกดปุ่มยุติ	
เกียรติศักดิ์ (2547)	244 ข้อ	Imax	Bayesian	SEE ≤ 0.30	
อำนาจ (2549)	$0 < a \leq +3$ $-3 < b < +3$ $c < 0.3$	b กับ $\hat{\theta}$	Bayesian	SEE < 0.45 หรือกดปุ่มยุติ	
สิริลักษณ์ (2549)	244 ข้อ $0.41 < a \leq 2.34$ $-2.23 < b < 2.72$ $0.06 < c < 0.30$	FI KL*	Bayesian	SEE ≤ 0.30* SEE ≤ 0.45	
ธงชัย (2551)	230 ข้อ $0.345 \leq a \leq 2.903$ $-3.640 \leq b \leq 3.839$ $0.082 \leq c \leq 0.300$	b กับ $\hat{\theta}$	Bayesian	SEE ≤ 0.30	
สุนันท์ (2551)	210 ข้อ $0.340 \leq a \leq 2.906$ $-3.640 \leq b \leq 3.839$ $0.074 \leq c \leq 0.300$	b กับ $\hat{\theta}$	Bayesian	SEE ≤ 0.30	

ผู้วิจัย	ประเด็นที่ศึกษา				
	Item Bank	Item Selection	Ability Estimate	Termination Criteria	Repeated answer
ธนศักดิ์ (2552)	103 ข้อ 0.51 < a ≤ 3.55 -1.87 < b < 2.87 0.0 < c < 0.29	Maximum Likelihood	MLE	Maximum Likelihood	
พิมพ์สิริ (2549)	210 ข้อ	Imax	Bayesian	20 ข้อ	ไม่ให้ทวน Block 5 ข้อ*
Urry (1977)	อย่างน้อย 100 ข้อ 0.8 < a -2.0 < b < 2.0 c < 0.3				
Weiss (1988)	118 – 150 ข้อ				
Hulin, drasgow and Parsons (1983)		b กับ $\hat{\theta}$ θ_{max} กับ $\hat{\theta}$ Imax*			
Chang and Ying (1996)		Global Information* Fisher			
Cheng, Ankenmann and Chang (2000)		FI, FII*, FIP* KL, KLP*			
Cheng and Ankenmann (2004)		FI FIP* KLP*, RN			

ผู้วิจัย	ประเด็นที่ศึกษา				
	Item Bank	Item Selection	Ability Estimate	Termination Criteria	Repeated answer
Weiss (1982)			MLE, Bayesian*		
Ho (1989)			MLE, Bayesian*		
Wang and Vispoel (1998)			MLE*, Bayesian* EAP*, Maximum a posteriore		
Dodd, De Ayala and Koch (1993)			SEE*, Test Information		
Gushta (2003)			25 ข้อ, SEE ≤ 0.30*		
Vispoel (1998)					ให้ทวน* ไม่ให้ทวน
Olea and others (2000)					ให้ทวน* ไม่ให้ทวน
Vispoel, Hendrickson and Bleiler (2000)					ไม่ให้ทวน Block 5 ข้อ* Block 10 ข้อ ทวนหลังสอบ
Revuelta, Ximenez and Olea (2003)					ไม่ให้ทวน ทวนหลังสอบ* Block 5 ข้อ * ทวนข้อต่อข้อ*

* คือ วิธีการที่ทดสอบแล้วมีประสิทธิภาพมากกว่า

ตาราง สรุปประเด็นในงานวิจัยที่เกี่ยวข้องกับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ

ผู้วิจัย	ประเด็นที่ศึกษา				
	Item Bank	Item Selection	Ability Estimate	Termination Criteria	Repeated answer
ทัศน์ศิรินทร์ (2554)	จำลอง 300 ข้อ ทุติยภูมิ 100 ข้อ ปฐมภูมิ 170 ข้อ	FI KL *	Bayesian	30 ข้อ SEE ≤ 0.3	
สมประสงค์ (2555)	คู่อันดับและกราฟ 59 ข้อ, สมการเชิงเส้นตัวแปรเดียว 104 ข้อ	Bayesian Volume Decrease	Bayesian	SEE < 0.049 SEE < 0.130 หรือ 15 ข้อ	
Van der Linden (1999)	50 ข้อ	Minimum error variance	Bayesian * MLE		
Cheng (2009a)	จำลอง 300 ข้อ จริง 36 ข้อ	KL SHE PWKL * HKL * RANDOM	MLE	จำลอง 12 ข้อ จริง 24 ข้อ	
Diao และ Reckase (2009)	50 ข้อ	FI * KL *	Bayesian * MLE	20 ข้อ 50 ข้อ *	
ณภัทร (2558)		MPI	MLE EAP		

* คือ วิธีการที่ทดสอบแล้วมีประสิทธิภาพมากกว่า