บทที่ 2
เอกสารและงานวิจัยที่เกี่ยวข้อง

งานวิจัยนี้ มีวัตถุประสงค์เพื่อพัฒนาแบบทดสอบวินิจฉัยทักษะการอ่านภาษาอังกฤษของนักศึกษาระดับปริญญาตรี โดยใช้โมเดลลำดับขั้นคุณลักษณะ พัฒนาวิธีการวินิจฉัยทักษะการอ่านภาษาอังกฤษ โดยใช้โมเดลลำดับขั้นคุณลักษณะและการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์สำหรับวินิจฉัยพื้นความรู้ของผู้สอบ (Knowledge States) ตามแบบแผนการตอบข้อสอบ โดยใช้วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ รวมทั้งเปรียบเทียบประสิทธิภาพของวิธีการวินิจฉัยระหว่างวิธีการทดสอบวินิจฉัยแบบปรับเหมาะด้วยคอมพิวเตอร์ กับวิธีการทดสอบวินิจฉัยแบบกระดาษดินสอ ด้านความคงที่ของจำนวนผู้สอบที่ตอบถูกในแต่ละคุณลักษณะ โดยผู้วิจัยได้ศึกษาแนวคิดจากเอกสารและงานวิจัยที่เกี่ยวข้องกับการวินิจฉัย การใช้โมเดลลำดับขั้นคุณลักษณะและการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ โดยนำเสนอเป็น 4 ตอน ดังนี้

ตอนที่ 1 แนวคิดเบื้องต้นเกี่ยวกับการวินิจฉัย
ตอนที่ 2 แนวคิดเบื้องต้นเกี่ยวกับโมเดลลำดับขั้นคุณลักษณะ
ตอนที่ 3 การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์
ตอนที่ 4 งานวิจัยเกี่ยวกับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์สำหรับการทดสอบวินิจฉัย

ตอนที่ 1 แนวคิดเกี่ยวกับการวินิจฉัย

ความหมายของการวินิจฉัย

การวินิจฉัยในความหมายโดยทั่วไปหมายถึง การตัดสิน ชี้ขาด การไตร่ตรอง หรือการใคร่ครวญ (ราชบัณฑิตยสถาน, 2554) ส่วนใหญ่ในทางการแพทย์ และทางพฤติกรรมศาสตร์ใช้เพื่อศึกษาอาการหรือพฤติกรรมเบี่ยงเบนที่เป็นปัญหาและค้นหาสาเหตุ เพื่อรักษาอาการหรือแก้ไขปรับปรุงพฤติกรรมเบี่ยงเบนนั้นๆ ส่วนทางการศึกษามีการใช้การวินิจฉัยอย่างแพร่หลายมากขึ้น ทั้งในบริบททางการศึกษาสำหรับบุคคลที่มีความต้องการพิเศษ เช่น การวินิจฉัยความบกพร่องของผู้เรียนที่มีความบกพร่องทางการเรียนรู้ (Learning Disability) โดยวินิจฉัยจากแบบแผนการตอบที่ผิดพลาดของบุคคลนั้นจากผู้เชี่ยวชาญ เป็นต้น และบริบททางการศึกษาสำหรับบุคคลทั่วไป มีนักการศึกษาและนักจิตวิทยาให้ความหมายการวินิจฉัยไว้หลายท่าน แต่ในที่นี้จะนำเสนอเฉพาะความหมายที่ให้โดยนักการศึกษาและนักจิตวิทยาในต่างประเทศ เนื่องจากในบริบทของการศึกษาในประเทศไทยพบเอกสารทางการวัดและประเมินผลการศึกษา ที่ให้ความหมายการวินิจฉัยไว้จำนวนน้อยมาก ส่วนใหญ่จะให้ความหมายในลักษณะเครื่องมือที่ใช้เพื่อการวินิจฉัยมากกว่า ความหมายของการวินิจฉัยที่ให้โดยนักการศึกษาและนักจิตวิทยาในต่างประเทศ สรุปได้ดังนี้

Delandshere (1990) ให้ความหมายของวิธีการวินิจฉัยไว้ว่าเป็นการทดสอบหรือการเปรียบเทียบลักษณะ (Profile) ของแต่ละบุคคลกับบรรทัดฐาน (Norms) หรือเกณฑ์ (Criteria) โดยเน้นการวินิจฉัยจุดเด่นและจุดด้อยของบุคคลที่เกี่ยวข้องกับกระบวนการของโรงเรียนเช่น หลักสูตร โปรแกรมการบริหารจัดการ

Hopkins and Antes (1990) ให้ความหมายของการวินิจฉัยไว้ว่า หมายถึงการจำแนกสภาวะที่ไม่ดีหรือไม่เป็นไปตามที่ต้องการในลำดับการเรียนรู้เพื่อปรับปรุงการเรียนการสอนให้ดีขึ้น

Cohen and Smith (1992) ให้ความหมายของการวินิจฉัยไว้ว่า หมายถึง การกระทำหรือกระบวนการในการจำแนกหรือกำหนดความผิดปกติ โดยการตรวจพิจารณาหรือการทดสอบ

Gregory (1992) ให้ความหมายของการวินิจฉัยไว้ว่า หมายถึง การกำหนดลักษณะที่เป็นอยู่หรือแหล่งของพฤติกรรมที่ไม่ปกติเพื่อจำแนกแบบแผนของพฤติกรรมตามระบบการวินิจฉัยที่ได้รับการยอมรับ

Gipps (1995) ให้ความหมายของการวินิจฉัยไว้ว่า หมายถึง การระบุปัญหาเฉพาะที่มีต่อความคิดหรือทักษะของผู้เรียนแต่ละบุคคล

Jonathan L. Goldman and Others (1996) ได้ให้ความหมายของการวินิจฉัยในพจนานุกรมของ Webster's new world ไว้ว่า หมายถึง การวิเคราะห์ถึงสาเหตุของสิ่งที่เป็นปัญหาในสภาวะใดสภาวะหนึ่ง หรือสถานการณ์ใดสถานการณ์หนึ่ง

Hornby and Sydney (2005) ได้ให้ความหมายของการวินิจฉัยในพจนานุกรมของอ๊อกฟอร์ด (Oxford) ไว้ว่า หมายถึง การค้นหาหรือการพิสูจน์เพื่อหาสาเหตุของสิ่งที่ทำให้เกิดสภาพที่ไม่ดีหรือสภาพที่เป็นปัญหา

Ketterlin and Yovanoff (2009) ได้ให้ความหมายของการวินิจฉัยไว้ว่า หมายถึง การประเมินผลที่ให้สารสนเทศเกี่ยวกับการเรียนรู้ของผู้เรียนในเรื่องความรู้และทักษะในขอบเขตที่กำหนดไว้ หรือมีความเข้าใจผิดเกี่ยวกับแนวคิดหรือเนื้อหาที่เรียนรู้

จากความหมายการวินิจฉัยข้างต้นสามารถวิเคราะห์เป็นข้อสรุปของการให้ความหมายของการวินิจฉัย จำแนกตามวิธีการที่ใช้ สิ่งที่วินิจฉัย เป้าหมายในการวินิจฉัย และประโยชน์ในการวินิจฉัยได้ดังนี้

1. วิธีการที่ใช้ ได้แก่ การระบุ การกำหนด การจำแนก การเปรียบเทียบ การค้นหา การทดสอบ การตรวจพิจารณา และการแปลความหมายข้อมูล ส่วนสิ่งที่ใช้ประกอบการดำเนินการอาจเป็นแบบแผนของพฤติกรรม บรรทัดฐาน (Norms) เกณฑ์ (Criteria) หรือระบบการวินิจฉัยที่ได้รับการยอมรับ

2. สิ่งที่วินิจฉัย ได้แก่ ปัญหา อุปสรรค ข้อบกพร่อง จุดเด่น จุดด้อย รวมทั้งสาเหตุที่ทำให้เกิดสิ่งต่าง ๆ เหล่านี้

3. เป้าหมายในการวินิจฉัย ได้แก่ ผู้เรียนและกระบวนการของสถานศึกษา ซึ่งเน้นไปที่แต่ละบุคคลมากกว่ากลุ่มบุคคล

4. ประโยชน์ในการวินิจฉัย ได้แก่ การปรับปรุงการจัดการเรียนการสอน

เมื่อพิจารณาความหมายของการวินิจฉัยดังกล่าวข้างต้น สรุปได้ว่า การวินิจฉัย หมายถึง การดำเนินการพัฒนาวิธีการวินิจฉัย หรือกระบวนการของการวินิจฉัย เพื่อค้นหาข้อบกพร่อง หรือจุดเด่น จุดด้อยในการเรียนรู้ของผู้เรียนแต่ละคน เกี่ยวกับกระบวนการคิด ความรู้ ทักษะ และความเข้าใจผิดในแนวคิดและเนื้อหาที่เรียน โดยใช้วิธีการตรวจพิจารณาหรือวิธีการทดสอบ โดยมีเป้าหมายเพื่อการปรับปรุงแก้ไข

ประเภทของการวินิจฉัย

Underhill (1972) ได้จำแนกประเภทของการวินิจฉัยตามระดับความละเอียดในการวินิจฉัยเป็น 3 ประเภท สรุปได้ดังนี้

1. การวินิจฉัยแบบทั่วไป (General Diagnosis) เป็นการวินิจฉัยระดับทั่วไป (General Level) ขั้นการสำรวจเพื่อทราบระดับความสามารถทั่ว ๆ ไป ของผู้เรียนทั้งรายบุคคลและรายกลุ่ม เครื่องมือที่ใช้ส่วนใหญ่เป็นแบบทดสอบทั้งแบบทดสอบวัดผลสัมฤทธิ์ที่ครูสร้างขึ้น และแบบทดสอบมาตรฐาน อาจใช้การวินิจฉัยโดยการเปรียบเทียบกับเกณฑ์ปกติ (Norm) ซึ่งจะทำให้ทราบข้อบกพร่องในด้านต่างๆ ของผู้เรียนอย่างกว้าง ๆ

2. การวินิจฉัยแบบวิเคราะห์ (Analytical Diagnosis) เป็นการวินิจฉัยระดับเฉพาะ (Specific Level) โดยการเก็บรวบรวมข้อมูลอย่างละเอียดเกี่ยวกับความสามารถของผู้เรียน เพื่อทราบข้อบกพร่องในตัวผู้เรียนเรื่องใดเรื่องหนึ่งโดยเฉพาะ เครื่องมือที่ใช้ส่วนใหญ่เป็นแบบทดสอบที่มีเนื้อหาเฉพาะเจาะจงในเรื่องนั้น ๆ การวินิจฉัยแบบวิเคราะห์มักดำเนินการภายหลังจากการวินิจฉัยแบบทั่วไปเพื่อให้ได้ข้อมูลที่มีรายละเอียดที่ลึกซึ้งมากขึ้น

3. การวินิจฉัยแบบคลินิก (Clinical Diagnosis) เป็นการวินิจฉัยที่ละเอียดลึกซึ้ง (Intensive Level) เพื่อให้เห็นถึงสาเหตุของปัญหาหรือข้อบกพร่องที่พบในตัวผู้เรียนซึ่งมีความซับซ้อน การใช้ข้อมูลจากแหล่งใดแหล่งหนึ่งเพียงแหล่งเดียวจึงไม่พอเพียง จำเป็นต้องใช้ข้อมูลจากหลายแหล่งประกอบกัน เช่น การใช้ข้อมูลจากแบบทดสอบ ประกอบการสัมภาษณ์ การสังเกต การศึกษาสภาพครอบครัว เป็นต้น

วิธีการที่ใช้ในการวินิจฉัย

ศิริเดช สุชีวะ (2538) ได้จำแนกรูปแบบของการวินิจฉัยทางการศึกษาไว้ 2 รูปแบบ ได้แก่ รูปแบบที่ไม่เป็นทางการและรูปแบบที่เป็นทางการ มีลักษณะและวิธีการวินิจฉัยที่ใช้ในแต่ละรูปแบบดังนี้

1. การวินิจฉัยที่ไม่เป็นทางการ (Informal Diagnosis) เป็นการค้นหาข้อบกพร่องของผู้เรียนด้วยวิธีการง่าย ๆ ไม่มีแบบแผนที่แน่นอน วิธีการที่ใช้ ได้แก่ การสังเกต การสอบถาม การสัมภาษณ์ การตรวจผลงาน ซึ่งวิธีการที่ใช้กันในระยะแรกคือ การสังเกตและการสอบถามผู้เรียนเป็นรายบุคคล วิธีนี้ผู้สอนต้องมีทักษะในการสังเกตและการตั้งคำถามเพื่อวินิจฉัย และใช้ได้ผลดีในกรณีที่เป็นผู้เรียนกลุ่มเล็ก และผู้สอนมีความใกล้ชิดกับผู้เรียน ยิ่งถ้าหากผู้สอนมีความชำนาญในเนื้อหาและการสอนด้วยแล้ว วิธีนี้จะให้ผลการวินิจฉัยที่ตรงกับความเป็นจริงสูง และยังคงมีการใช้กันอยู่ในปัจจุบัน แต่มีข้อจำกัดในเรื่องของเวลา เพราะต้องใช้เวลาในการวินิจฉัยค่อนข้างมาก รวมทั้งจำนวนผู้เรียนและจำนวนห้องเรียนที่ผู้สอนต้องรับผิดชอบมีมากขึ้น ทำให้ไม่สะดวกสำหรับผู้สอน ด้วยความจำกัดของเวลาที่กำหนดไว้สำหรับการเรียนในแต่ละเนื้อหาที่มีมาก และเวลาที่ผู้สอนต้องปฏิบัติภาระงานอื่นที่ได้รับมอบหมาย นอกเหนือจากภาระงานสอน (ศิริเดช สุชีวะ, 2538) ส่วนวิธีการวินิจฉัยจากการให้ผู้เรียนทำแบบฝึกหัด หรือแบบทดสอบในชั้นเรียนหลังเรียนจบในเนื้อหานั้น ๆ ซึ่งใช้ได้ผลดีทั้งในการวินิจฉัยมโนทัศน์พื้นฐานและมโนทัศน์ที่ซับซ้อน แต่ผู้สอนที่ทำการวินิจฉัยจะต้องเชี่ยวชาญในเนื้อหานั้นเป็นอย่างดี และเวลาที่ใช้ในการสอบ การตรวจและการวินิจฉัยใช้เวลาค่อนข้างมากทำให้เป็นข้อจำกัดสำคัญของวิธีการนี้ เนื่องจากสภาพการเรียนการสอนปัจจุบันผู้สอนแทบไม่มีเวลาที่จะ

ตรวจงานของผู้เรียนด้วยตนเองเป็นรายบุคคล และถ้าหากผู้เรียนลอกแบบฝึกหัดที่มอบหมายให้ทำมาส่ง ผู้สอนจะยิ่งไม่สามารถวินิจฉัยมโนทัศน์ที่คลาดเคลื่อนจากการบ้านหรือแบบฝึกหัดได้เลย (ศิริเดช สุชีวะ, 2538)

2. การวินิจฉัยที่เป็นทางการ (Formal Diagnosis) เป็นการค้นหาข้อบกพร่องของผู้เรียนโดยวิธีการที่สร้างขึ้นมาอย่างเป็นระบบและมีแบบแผน โดย การใช้แบบทดสอบวินิจฉัยหรือการวินิจฉัยจากแบบแผนการตอบข้อสอบ โดยใช้ดัชนีบ่งชี้ความผิดปกติของแบบแผนการตอบข้อสอบ การใช้วิธีการ Rule space หรือการใช้การย้อนรอยกระบวนการคิด เป็นต้น ทั้งนี้การวินิจฉัยแบบเป็นทางการที่นิยมใช้กันมาเป็นเวลายาวนานในวงการวัดผลการศึกษา คือ การใช้แบบทดสอบวินิจฉัย (ศิริเดช สุชีวะ, 2538)

ตอนที่ 2 แนวคิดเกี่ยวกับโมเดลลำดับขั้นคุณลักษณะ

โมเดลลำดับขั้นคุณลักษณะ (Attribute Hierarchy Model: AHM) เป็นโมเดลทฤษฎีการตอบสนองข้อสอบทางปัญญา (Cognitive Item Response Theory Model) ที่พัฒนาขึ้นครั้งแรกโดย Leighton, Gierl and Hunka ในปี ค.ศ. 2000 จากศูนย์วิจัยการวัดประเมินผลการศึกษา มหาวิทยาลัยอัลเบอร์ต้า (Center for Research in Applied Measurement and Evaluation, University of Alberta) และได้นำเสนอแนวคิดนี้ไว้ในวารสารการวัดผลการศึกษาของสภาการวัดและประเมินผลการศึกษาของสภาแห่งชาติ (National Council on Measurement in Education: NCME) ในปี ค.ศ. 2000 โดยใช้ชื่อเรื่องว่าการวิเคราะห์และพัฒนาแบบทดสอบวินิจฉัยทางตรรกวิทยาโดยใช้รูลสเปซโมเดลของทาสซูโอกะ (Exploring the Logic of Tatsuoka's Rule Space Model for Test Development and Analysis) และได้นำเสนอในงานประชุมสัมมนาการวัดและประเมินผลการศึกษาของสภาแห่งชาติ (National Council on Measurement in Education: NCME) ในชื่อเรื่องโมเดลลำดับขั้นคุณลักษณะสำหรับการประเมินทางพุทธิปัญญา (The Attribute Hierarchy Model for Cognitive Assessment) ต่อมาได้รับการตีพิมพ์ในวารสารการวัดผลการศึกษา (Journal of Educational Measurement) ในปี ค.ศ. 2004 โดยใช้ชื่อเรื่องว่า โมเดลลำดับขั้นคุณลักษณะสำหรับการประเมินทางพุทธิปัญญากรณีศึกษาความ ผันแปรโดยวิธีรูลสเปซของทาสซูโอกะ (The Attribute Hierarchy Model for Cognitive Assessment: A Variation on Tatsuoka's Rule Space Approach)

โมเดลลำดับขั้นคุณลักษณะ (Attribute Hierarchy Model: AHM) เป็นโมเดลทฤษฎีการตอบสนองข้อสอบทางปัญญา (Cognitive Item Response Theory Model) ที่พัฒนาขึ้นเพื่อแก้ไขข้อจำกัดของรูลสเปซโมเดล (Rule Space Model) เพื่อให้การวินิจฉัยมีความชัดเจนและเป็นประโยชน์ต่อการพัฒนาการเรียนการสอน และพัฒนาศักยภาพของผู้เรียนมากขึ้น โดยมีการกำหนดจำนวนข้อสอบ ลักษณะข้อสอบและแบบแผนการตอบข้อสอบจากโครงสร้างที่เป็นลำดับขั้นของความรู้หรือทักษะที่ต้องใช้ในการตอบข้อสอบซึ่งเรียกว่าคุณลักษณะ (Attribute) ซึ่งแสดงถึงการทำงานที่เชื่อมโยงกันระหว่างทฤษฎีทางปัญญา (Cognitive Theory) กับการวัดทางจิตวิทยา (Psychometric) ซึ่งเป็นโมเดลที่มีลักษณะและขั้นตอนที่สำคัญของการนำไปใช้ในการพัฒนาแบบทดสอบวินิจฉัย ดังนี้

1. ลักษณะสำคัญของโมเดลลำดับขั้นคุณลักษณะ

ลักษณะสำคัญของโมเดลลำดับขั้นคุณลักษณะคือ การให้คำจำกัดความและการจำแนกคุณลักษณะ รวมทั้งการกำหนดลักษณะความสัมพันธ์ที่เป็นลำดับขั้น ดังนี้

1.1 คุณลักษณะเป็นตัวแปรนำเข้าที่สำคัญในรูลสเปซโมเดล (Rule Space Model) เพราะเป็นการสรุปอ้างอิงถึงทักษะทางปัญญาของผู้สอบ แต่จากเอกสารงานวิจัยที่เกี่ยวข้องของการนำรูลสเปซโมเดล (Rule Space Model) ไปใช้ ไม่ได้ให้คำจำกัดความคุณลักษณะไว้อย่างชัดเจน แต่มีการให้คำจำกัดความที่แตกต่างกันออกไป รวมทั้งมีการใช้เทคนิคในการจำแนกที่หลากหลาย นักวิจัยบางคน กล่าวถึงคุณลักษณะในความหมายทั่ว ๆ ไป โดยพิจารณาจากทักษะทางปัญญาที่จำเป็นต้องใช้ในการปฏิบัติงาน บางคนกล่าวถึงคุณลักษณะในความหมายเฉพาะในลักษณะของการรวมองค์ประกอบย่อยทางปัญญาเพื่อใช้ในการแก้ปัญหา ทักษะพื้นฐานทางปัญญาที่ต้องใช้เพื่อความรอบรู้ในเนื้อหาเฉพาะหรือชุดของวิธีการหรือการปฏิบัติที่ใช้ในการแก้ปัญหาในเนื้อหาเฉพาะคุณลักษณะจึงถูกนำไปใช้ทั้งในความหมายเฉพาะและความหมายทั่วไป (Gierl, Leighton & Hunka, 2007) นอกจากนี้ Birenbaum and Tatsuoka (1993) ยังมีการให้คำจำกัดความเกี่ยวกับคุณลักษณะไว้ว่า "เป็นการบรรยายกระบวนการ ทักษะหรือความรู้ของผู้เรียนที่ต้องใช้เพื่อให้สามารถทำงานได้บรรลุเป้าหมาย แต่ไม่ได้จำกัดเฉพาะความสามารถของผู้เรียน ในบางวิธีการคุณลักษณะอาจรวมถึงสิ่งที่ผู้เรียนใช้เป็นยุทธวิธีเพื่อช่วยในการเรียนรู้" ซึ่ง Gierl, Leighton and Hunka (2007) ได้ยกตัวอย่างวิธีการที่ใช้ในการจำแนกคุณลักษณะไว้สามารถสรุปได้ ดังนี้

1.1.1 การใช้กรอบมโนทัศน์โดยการใช้โมเดลการแก้ปัญหาในการประเมินแบบทดสอบทางสถาปัตยกรรมจำนวน 22 ข้อ โมเดลมี 3 ขั้นตอนได้แก่ สร้างตัวแทนของข้อสอบขั้นต้น กำหนดเป้าหมายและการปฏิบัติในแต่ละเป้าหมายทั้งที่สำเร็จและที่พึงพอใจ คุณลักษณะของข้อสอบได้มาจากกระบวนการทางปัญญาที่มีลักษณะเฉพาะในแต่ละขั้นของโมเดล โดย Sheehan and Tatsuoka (1998) หรือการใช้กรอบมโนทัศน์ในลักษณะทั่วไปในการจัดกลุ่มข้อสอบและโมเดลทางปัญญาโดย Sheehan, Tatsuoka and Lawis (1992)

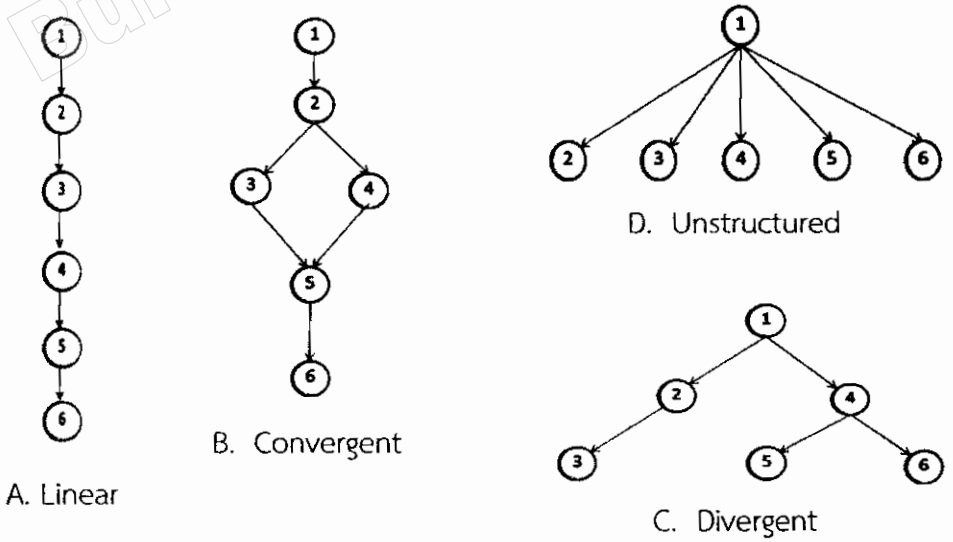
1.1.2 ใช้การจับคู่ประโยคและวิธีการที่ใช้ในทฤษฎีฟาเซท ต่อการจำแนกคุณลักษณะที่เป็นเนื้อหาและกระบวนการจากข้อมูลต้นร่าง (Protocol Data) ของ Scholastic Assessment Test (SAT) โดยจับคู่ประโยคได้ 13 ฟาเซท แต่ละฟาเซทจะประกอบด้วยความรู้เบื้องต้น (Element) เพื่อนำมาจำแนกเป็นคุณลักษณะสำหรับแบบทดสอบ

1.1.3 ใช้การตัดสินของผู้เชี่ยวชาญ ซึ่ง Tatsuoka ได้ให้ข้อเสนอแนะไว้ว่าการวิเคราะห์งานขั้นต้นนั้นสามารถให้ผู้เชี่ยวชาญหรือครูที่เชี่ยวชาญหลายๆ คน ถ้าผู้เชี่ยวชาญจำนวน 2 คน หรือมากกว่ามีวิธีการแก้ปัญหาที่ต่างกันสามารถใช้คุณลักษณะที่ต่างกันในเมตริกซ์ของข้อสอบได้ ในกรณีนี้อาจจะมี Incidence Matrices ที่ต่างกันหลายเมตริกซ์ และแต่ละเมตริกซ์สามารถทดสอบโดย รูลสเปซโมเดล (Rule Space Model) ได้ ทั้งนี้อาจจำแนกคุณลักษณะจากวิธีการหรือใช้แหล่งข้อมูลที่ต่างกัน เช่น ใช้ผู้เชี่ยวชาญด้านเนื้อหาหรือจากเอกสารต้นร่างของผู้เรียน เป็นต้น

ความไม่ชัดเจนในการให้คำจำกัดความและการจำแนกคุณลักษณะดังกล่าว เป็นเหตุให้คุณลักษณะที่ปรากฏในอกสารที่เกี่ยวข้องกับการนำรูลสเปซโมเดล (Rule Space Model) ไปใช้มี

ข้อจำกัดด้านมโนทัศน์เบื้องต้น เนื่องจากการใช้เทคนิค โมเดลหรือวิธีการที่แตกต่างกันทำให้คุณลักษณะมีคุณภาพที่แตกต่างกัน จึงไม่สามารถเชื่อมโยงพื้นฐานความรู้จากการศึกษาหนึ่งไปยังการศึกษาอื่น ๆ ได้ โดยเฉพาะประโยชน์ในการแปลความหมายจากผลคะแนนที่ได้จากการทดสอบ Gierl, Leighton and Hunka (2007) ได้กล่าวถึงการแก้ไขข้อจำกัดเกี่ยวกับคุณลักษณะว่า "คุณลักษณะเป็นการบรรยายวิธีการหรืออธิบายความรู้ที่ต้องใช้ในการทำงานที่มีขอบเขตเฉพาะ แม้ว่าคุณลักษณะจะไม่ใช้ยุทธวิธี แต่จะเป็นการช่วยสร้างยุทธวิธี ยิ่งไปกว่านั้น ชุดของคุณลักษณะจะถูกจัดเข้าสู่ยุทธวิธีเพื่อสนับสนุนบทบาทในการแก้ปัญหาที่เกิดขึ้นทันทีทันใด โดยไม่จำเป็นต้องเป็นกลุ่มยุทธวิธี คุณลักษณะเป็นเอกลักษณ์ที่มีลักษณะเป็นพลวัต ชุดของคุณลักษณะของผู้เรียนที่มีความสามารถในเวลาหนึ่งอาจไม่เหมือนชุดของคุณลักษณะของผู้เรียนที่มีความสามารถในอีกเวลาหนึ่ง อาจเป็นพัฒนาการในตัวผู้เรียนเอง หรือขึ้นอยู่กับการจัดการเรียนการสอน นั่นคือ ความก้าวหน้าของผู้เรียนจากเวลาหนึ่งไปยังอีกเวลาหนึ่งอาจเกิดขึ้นจากการมีพัฒนาการที่เพิ่มขึ้น หรือองค์ประกอบของการจัดการเรียนการสอน คุณลักษณะสำหรับแบบทดสอบสามารถจำแนกได้โดยใช้วิธีการที่แตกต่างกัน เช่น จากความคิดเห็นของผู้เชี่ยวชาญในเนื้อหาวิชานั้น ๆ จากการวิเคราะห์งาน การเขียนตอบของผู้เรียน โดยควรทำการตรวจสอบความตรงของการจำแนกคุณลักษณะทั้งจากตัวของผู้สอบและข้อสอบเปรียบเทียบกับกลุ่มเป้าหมาย"

1.2 การกำหนดลักษณะความสัมพันธ์ที่เป็นลำดับขั้นคุณลักษณะ นอกจากการให้คำจำกัดความและการจำแนกคุณลักษณะได้อย่างถูกต้องแล้ว สิ่งสำคัญประการหนึ่งในการนำโมเดลลำดับขั้นของคุณลักษณะไปใช้ได้อย่างมีประสิทธิภาพคือ การกำหนดลักษณะความสัมพันธ์ที่เป็นลำดับขั้นที่ถูกต้อง โดยอาจพิจารณาเชิงประจักษ์จากการให้คำจำกัดความที่ชัดเจนหรือจากการวิเคราะห์ต้นร่างเอกสารหรือพิจารณาเชิงทฤษฎีจากจิตวิทยาพัฒนาการ เช่น จากขั้นพัฒนาการของเพียเจต์ เป็นต้น สำหรับโครงสร้างลำดับขั้นของ Leighton Gierl and Hunka (2007) เป็นตัวอย่างโครงสร้างลำดับขั้นอย่างง่ายที่ใช้ในการพัฒนาแบบทดสอบมี 4 รูปแบบ ดังภาพที่ 4



ภาพที่ 4 โครงสร้างลำดับขั้นที่ใช้ในการพัฒนาแบบทดสอบ 4 รูปแบบของ Leighton, Gierl and Hunka (2002)

1.2.1 ลำดับขั้นเชิงเส้น (Linear Hierarchy) มีลักษณะความสัมพันธ์ระหว่างคุณลักษณะในลักษณะเส้นตรงที่มีจุดสิ้นสุดจุดเดียว (ดังภาพที่ 3-A) ซึ่งแสดงให้เห็นว่า คุณลักษณะที่ 1 จะต้องมีมาก่อนคุณลักษณะที่ 2 คุณลักษณะที่ 2 จะต้องมีมาก่อนคุณลักษณะที่ 3 คุณลักษณะที่ 3 จะต้องมีมาก่อนคุณลักษณะที่ 4 คุณลักษณะที่ 4 จะต้องมีมาก่อนคุณลักษณะ ที่ 5 และคุณลักษณะ ที่ 5 จะต้องมีมาก่อนคุณลักษณะที่ 6

1.2.2 ลำดับขั้นเชิงลู่เข้า (Hierarchy with a Convergent Branch) มีลักษณะความสัมพันธ์ที่แยกเป็น 2 ทางแต่มีจุดสิ้นสุดจุดเดียว (ดังภาพที่ 3-B) ซึ่งแสดงให้เห็นว่า คุณลักษณะที่ 1 จะต้องมีมาก่อนคุณลักษณะที่ 2 คุณลักษณะที่ 2 จะต้องมีมาก่อนคุณลักษณะที่ 3 และคุณลักษณะที่ 4 คุณลักษณะที่ 3 และคุณลักษณะที่ 4 จะต้องมีมาก่อนคุณลักษณะที่ 5 และคุณลักษณะที่ 5 จะต้องมีมาก่อนคุณลักษณะที่ 6

1.2.3 ลำดับขั้นเชิงจำแนก (Hierarchy having a Divergent Branch) มีลักษณะความสัมพันธ์ที่แยกออกเป็น 2 ทาง แต่ไม่ได้มีจุดสิ้นสุดเพียงจุดเดียว (ดังภาพที่ 3-C) ซึ่งแสดงให้เห็นว่า คุณลักษณะที่ 1 จะต้องมีมาก่อนคุณลักษณะที่ 2 และคุณลักษณะที่ 4 คุณลักษณะที่ 2 และคุณลักษณะที่ 3 คุณลักษณะที่ 4 จะต้องมีมาก่อนคุณลักษณะที่ 5 และคุณลักษณะที่ 6

1.2.4 ลำดับขั้นแบบไม่มีโครงสร้าง (Unstructured Hierarchy) มีลักษณะความสัมพันธ์ที่ไม่มีความสัมพันธ์เดียวจากจุดเริ่มต้นและไม่มีจุดสิ้นสุดเพียงจุดเดียว (ดังภาพที่ 3-D) ซึ่งแสดงให้เห็นว่า คุณลักษณะที่ 1 จะต้องมีมาก่อนคุณลักษณะที่ 2 คุณลักษณะที่ 3 คุณลักษณะที่ 4 คุณลักษณะที่ 5 และคุณลักษณะที่ 6 โดยที่คุณลักษณะที่ 2 ถึงคุณลักษณะที่ 6 ไม่มีความสัมพันธ์กัน

2. ขั้นตอนการนำโมเดลลำดับขั้นคุณลักษณะไปใช้ในการพัฒนาแบบทดสอบวินิจฉัย

การนำโมเดลลำดับขั้นคุณลักษณะไปใช้ในการพัฒนาแบบทดสอบและการวินิจฉัยประกอบด้วยขั้นตอนสำคัญ 7 ขั้นตอนดังต่อไปนี้

2.1 ระบุลักษณะและความสัมพันธ์เชิงลำดับขั้นที่มีลักษณะเฉพาะ เป็นการทำความเข้าใจกระบวนการทางปัญญาที่จะใช้วัดกลุ่มผู้สอบที่มีลักษณะเฉพาะให้ชัดเจน ให้คำจำกัดความคุณลักษณะอย่างรอบคอบ กำหนดความสัมพันธ์ที่เป็นลำดับขั้นและนำเสนอลำดับขั้นโดยใช้โครงสร้างรูปต้นไม้ (Tree Structure) ความสัมพันธ์ที่เป็นลำดับขั้นในข้อสอบแต่ละข้อจะต้องนำเสนอเพื่อให้ผู้สอบไม่สับสนในการตอบข้อสอบ แม้ว่าจะมีปฏิสัมพันธ์กันระหว่างคุณลักษณะที่ทำได้กับลำดับที่นำเสนอให้ทำภายในเฉพาะแต่ละข้อ เช่น ยุทธวิธีในการนำเสนอจะต้องไม่มี 2 มโนทัศน์ในแต่ละข้อต้องวิเคราะห์และระบุลักษณะทางปัญญาอย่างชัดเจน สมเหตุสมผล และเป็นไปตามหลักจิตวิทยา ทั้งนี้ควรกำหนดคุณลักษณะที่เป็นจุดเริ่มต้นซึ่งเป็นสิ่งที่ต้องมีมาก่อนสำหรับคุณลักษณะทั้งหมดจากจุดเดียวกัน เพื่อช่วยให้นักพัฒนาแบบทดสอบตระหนักถึงคุณลักษณะเริ่มต้นที่ผู้สอบทุกคนต้องมี

2.2 กำหนดเมตริกซ์เพื่อใช้ในการพัฒนาแบบทดสอบ ประกอบด้วยขั้นตอนดังต่อไปนี้

2.2.1 กำหนด Binary Adjacency Matrix A เพื่อแสดงความสัมพันธ์ทางตรงระหว่างคุณลักษณะ โดยกำหนดให้ Matrix A มีขนาด k X k เมื่อ k เป็นจำนวนคุณลักษณะเลข 1 ณ ตำแหน่ง (j, k) ในเมตริกซ์จะบ่งบอกถึงการมีความสัมพันธ์ทางตรงระหว่างคุณลักษณะที่ j ในลักษณะของการที่เป็นคุณลักษณะที่ต้องมีมาก่อนคุณลักษณะที่ k ซึ่งสามารถเขียนได้ ดังนี้

$$\begin{bmatrix}010100\\001000\\000000\\000011\\000000\\000000\end{bmatrix}$$

จากเมตริกซ์แสดงให้เห็นว่าคุณลักษณะที่ 1 เป็นคุณลักษณะที่ต้องมีมาก่อนคุณลักษณะที่ 2 และคุณลักษณะที่ 4 เห็นได้จากตำแหน่งที่ (1,2) และ (1,4) แทนด้วยเลข 1 คุณลักษณะที่ 2 เป็นคุณลักษณะที่ต้องมีมาก่อนคุณลักษณะที่ 3 เห็นได้จากตำแหน่งที่ (2,3) แทนด้วยตัวเลข 1 และคุณลักษณะที่ 4 เป็นคุณลักษณะที่ต้องมีมาก่อนคุณลักษณะที่ 5 และ 6 เห็นได้จากตำแหน่งที่ (4,5) และ (4,6) แทนด้วยเลข 1

2.2.2 กำหนด Reach Ability Matrix R เพื่อแสดงความสัมพันธ์ทางตรงและความสัมพันธ์ทางอ้อมระหว่างคุณลักษณะ โดยกำหนดให้ Matrix R มีขนาด k X k เมื่อ k เป็นจำนวนคุณลักษณะ Matrix R จะบ่งบอกเงื่อนไขในโครงสร้างของลำดับขั้นซึ่งเป็นประโยชน์ในการกำหนดชุดย่อยของข้อคำถาม คำนวณโดยใช้สูตร $R = (A+I)^n$ เมื่อ n เป็นจำนวนเต็มและ I เป็นเมตริกซ์เอกลักษณ์ (Identity Matrix) จาก Matrix A ในข้อที่ 2.2.1 สามารถกำหนดเป็น Matrix R ได้ดังนี้

$$\begin{bmatrix}111111\\011000\\001000\\000111\\000010\\000001\end{bmatrix}$$

แถวที่ 1 แสดงให้เห็นว่าคุณลักษณะที่ 1 เป็นคุณลักษณะที่ต้องมีมาก่อนทุกคุณลักษณะ

แถวที่ 2 แสดงให้เห็นว่า คุณลักษณะที่ 2 เป็นคุณลักษณะที่ต้องมีมาก่อน คุณลักษณะที่ 3

แถวที่ 3 แสดงให้เห็นว่า คุณลักษณะที่ 3 ไม่ได้เป็นคุณลักษณะที่ต้องมีมาก่อนคุณลักษณะใด ๆ

แถวที่ 4 แสดงให้เห็นว่า คุณลักษณะที่ 4 เป็นคุณลักษณะที่ต้องมีมาก่อนคุณลักษณะที่ 5 และ 6

แถวที่ 5 และแถวที่ 6 แสดงให้เห็นว่า คุณลักษณะที่ 5 และคุณลักษณะที่ 6 ไม่ได้เป็นคุณลักษณะที่ต้องมีมาก่อนคุณลักษณะใด

2.2.3 กำหนด Incidence Matrix Q เพื่อแสดงชุดของข้อสอบและคุณลักษณะที่ต้องใช้ในการทำข้อสอบในแต่ละข้อให้ถูกต้อง ชุดของข้อสอบที่เป็นไปได้มีขนาดเท่ากับ 2^k-1 เมื่อ k เป็นจำนวนคุณลักษณะ Matrix Q มีขนาดเท่ากับ (k,i) เมื่อ k เป็นจำนวนคุณลักษณะ และ i เป็นจำนวนข้อสอบที่เป็นไปได้ทั้งหมดจาก Matrix R ในข้อที่ 2 สามารถกำหนดเป็น Matrix Q ได้ดังนี้

มหาวิทยาลัยบูรพา Burapha University

$$\begin{bmatrix}
101010101010101010101010101010101010101010101010101010101010101 \\
011001100110011001100110011001100110011001100110011001100110011 \\
000111100001111000011110000111100001111000011110000111100001111 \\
000000011111111000000001111111100000000111111110000000011111111 \\
000000000000000011111111111111110000000000000000111111111111111 \\
000000000000000000000000000000001111111111111111111111111111111
\end{bmatrix}$$

จากเมตริกซ์แสดงให้เห็นว่าชุดของข้อสอบที่เป็นไปได้มีจำนวน 63 แบบ (2^6-1) และมีคุณลักษณะที่ต้องใช้ในการตอบข้อสอบให้ถูกต้องจำนวน 1 ถึง 6 คุณลักษณะ ในแต่ละข้อจะมีจำนวนและคุณลักษณะที่ใช้ในการตอบข้อสอบแตกต่างกันไป เช่น ในคอลัมน์ที่ 1 แสดงถึงข้อสอบชุดที่ 1 ต้องใช้คุณลักษณะที่ 1 เพียงคุณลักษณะเดียวในการตอบข้อสอบให้ถูกต้อง ในคอลัมน์ที่ 2 แสดงถึงข้อสอบชุดที่ 2 ที่ต้องใช้คุณลักษณะที่ 2 เพียงคุณลักษณะเดียวในการตอบข้อสอบให้ถูกต้อง ในคอลัมน์ที่ 3 แสดงถึงข้อสอบชุดที่ 3 ที่ต้องใช้คุณลักษณะที่ 1 และ 2 จึงจะทำข้อสอบได้ถูกต้อง

2.2.4 กำหนด Reduced Matrix Q เพื่อลดจำนวนชุดของข้อสอบที่เป็นไปได้ตามเงื่อนไขโครงสร้างลำดับขั้นที่กำหนดจาก Matrix Q แสดงให้เห็นว่าชุดของข้อสอบที่เป็นไปได้จะมีจำนวนมากหรือน้อยขึ้นอยู่กับจำนวนคุณลักษณะซึ่งจะเห็นได้ว่า ถึงแม้ว่าคุณลักษณะจะมีจำนวนไม่มากแต่ชุดของข้อสอบที่เป็นไปได้มีจำนวนมาก เช่น จากตัวอย่างในข้อที่ 2.2.3 จะเห็นได้ว่ามีจำนวนคุณลักษณะเพียง 6 คุณลักษณะ แต่มีข้อสอบที่เป็นไปได้ถึง 63 แบบ ดังนั้นเพื่อให้ข้อสอบมีจำนวนลดลงและมีลักษณะข้อสอบที่มีความเฉพาะมากขึ้น ในการกำหนดชุดของข้อสอบที่เป็นไปได้จึงสามารถทำได้โดยการนำเงื่อนไขโครงสร้างของลำดับขั้นมาพิจารณาและตัดชุดข้อสอบที่ไม่เป็นไปตามเงื่อนไขออกไป เช่น จากตัวอย่าง Matrix Q ในข้อที่ 2.2.3 สามารถลดจำนวนข้อสอบตามเงื่อนไขโครงสร้างของลำดับขั้นเหลือเพียง 15 แบบ

$$\begin{bmatrix}
111111111111111 \\
011011011011011 \\
001001001001001 \\
000111111111111 \\
000000111000111 \\
000000000111111
\end{bmatrix}$$

2.3 พัฒนาแบบทดสอบโดยดำเนินการสร้างข้อสอบให้มีการใช้คุณลักษณะตามที่กำหนด

2.4 กำหนดแบบแผนการตอบข้อสอบที่คาดหวัง (Expected Item Response Pattern) คะแนนรวม (Total Score) และคุณลักษณะของผู้สอบในแต่ละแบบแผนการตอบข้อสอบ

2.5 นำแบบทดสอบไปใช้

2.6 ประมาณค่าความน่าจะเป็นในการตอบข้อสอบ (Estimating Probability of Item Response) โดยทำการประมาณค่าความน่าจะเป็นในการตอบข้อสอบในแต่ละข้อ จากแบบแผนการตอบข้อสอบที่คาดหวัง โดยใช้ดัชนีความสอดคล้องกับบุคคล (Person-fit Indices) จาก

โอกาสในการตอบข้อสอบได้ถูกต้อง โดยใช้โมเดลทฤษฎีการตอบสนองข้อสอบ (IRT Model) เพื่อให้ได้ค่าอำนาจจำแนกของข้อสอบ (a-Parameter) ค่าความยากของข้อสอบ (b-Parameter) และค่าการเดาข้อสอบ (c-Parameter) โดยใช้โปรแกรมสำเร็จรูปช่วยในการคำนวณ

2.7 จำแนกแบบแผนการตอบข้อสอบที่สังเกตได้ ซึ่งประกอบด้วยขั้นตอนดังต่อไปนี้

2.7.1 การจำแนกเบื้องต้น (Preliminary Classification) โดยใช้วิธีการเปรียบเทียบแบบแผนการตอบข้อสอบที่สังเกตได้ กับแบบแผนการตอบข้อสอบที่คาดหวังทั้งหมด และคำนวณความเป็นไปได้สูงสุดที่แบบแผนการตอบข้อสอบที่สังเกตได้ สอดคล้องกับแบบแผนการตอบข้อสอบที่คาดหวัง ถ้ามีความเป็นไปได้สูงสุดที่แบบแผนการตอบที่คาดหวังใด แสดงว่าแบบแผนการตอบข้อสอบที่สังเกตได้นั้น มีคุณลักษณะของผู้สอบตามแบบแผนการตอบข้อสอบที่คาดหวังนั้นกำหนด ตัวอย่างเช่น นางสาว ก. มีแบบแผนที่สังเกตได้เป็น 111100000000000 เมื่อนำไปเปรียบเทียบคำนวณหาค่าความน่าจะเป็นในแต่ละแบบแผนการตอบข้อสอบที่คาดหวัง พบว่า มีความเป็นไปได้สูงสุดเท่ากับ 0.50 ที่ระดับความสามารถเท่ากับ -0.50 ตรงกับแบบแผนการตอบข้อสอบที่คาดหวัง 111000000000000 ซึ่งมีความคลาดเคลื่อนเกิดขึ้น 1 จุด และมีคุณลักษณะของผู้สอบเป็น 111000 แสดงว่า นางสาว ก. จัดให้อยู่ในกลุ่มที่มีคุณลักษณะ ที่ 1, 2 และ 3 เป็นต้น แบบแผนการตอบข้อสอบที่สังเกตได้ในการจำแนกเบื้องต้น มีรายละเอียดการคำนวณและสูตรที่ใช้ในการคำนวณดังนี้

2.7.1.1 คำนวณโอกาสในการเกิดความคลาดเคลื่อนจาก $1 \rightarrow 0$ และจาก $0 \rightarrow 1$ เพื่อคำนวณความเป็นไปได้สูงสุดที่แบบแผนการตอบข้อสอบที่สังเกตได้จะสอดคล้องกับแบบแผนการตอบข้อสอบที่คาดหวัง โดยกำหนดให้

V_j เป็นแบบแผนการตอบข้อสอบที่คาดหวังที่ j สำหรับข้อสอบ n ข้อ

X เป็นแบบแผนการตอบข้อสอบที่สังเกตได้ซึ่งมีจำนวนข้อเท่ากับ V_j

$D_j = V_j - X$ เป็นความแตกต่างระหว่างแบบแผนการตอบข้อสอบที่คาดหวังกับแบบแผนการตอบข้อสอบที่สังเกตได้

ถ้า D = 0 แสดงว่าไม่มีความคลาดเคลื่อน

ถ้า D = -1 แสดงว่ามีความคลาดเคลื่อนจาก $0 \rightarrow 1$

ถ้า D = 1 แสดงว่ามีความคลาดเคลื่อนจาก $1 \rightarrow 0$

2.7.1.2 คำนวณค่าความน่าจะเป็นที่เป็นแบบแผนการตอบข้อสอบที่สังเกตได้จะจัดอยู่ในแบบแผนการตอบข้อสอบที่คาดหวัง โดยการรวมโอกาสที่จะตอบถูกเมื่อผิดและโอกาสที่จะตอบผิดเมื่อถูก ด้วยหลักการของ Boolean Algebra ตามสูตรดังต่อไปนี้

กำหนดให้ $P_{jk}(\theta)$ เป็นโอกาสที่จะตอบถูกเมื่อผิด $(0 \rightarrow 1)$

$1 - P_{jk}(\theta)$ เป็นโอกาสที่จะตอบผิดเมื่อถูก $(1 \rightarrow 0)$

$$P_{jExpected}(\theta) = \prod_{k=1}^{K} P_{jk}(\theta) \prod_{m=1}^{M} \left[1 - P_{jm}(\theta)\right]$$

2.7.1.3 การตรวจสอบการจำแนกเบื้องต้น (Verification of Preliminary Classification) การจำแนกเบื้องต้นบางครั้งให้ค่าความน่าจะเป็นต่ำมากและจำแนกได้ไม่สอดคล้องกับโครงสร้างของลำดับขั้นเช่น นางสาว ก. มีแบบแผนการตอบที่สังเกตได้เป็น 111111111111101 เมื่อคำนวณค่าความน่าจะเป็นกลับพบว่า มีค่าความเป็นไปได้สูงสุดเท่ากับ 0.0318 ที่ระดับความสามารถเท่ากับ 2.37 ตรงกับแบบแผนการตอบข้อสอบที่คาดหวัง คือ 111111111111111 ซึ่งมีความคลาดเคลื่อนเกิดขึ้น 1 จุด และมีคุณลักษณะของผู้สอบเป็น 111111 แสดงว่านางสาว ก. ถูกจัดให้อยู่ในกลุ่มที่มีคุณลักษณะครบทุกข้อ ซึ่งในความเป็นจริงไม่ได้เป็นเช่นนั้น เนื่องจากนางสาว ก. ตอบข้อสอบได้ไม่ถูกต้องทุกข้อ ทำให้การจำแนกไม่ตรงกับข้อมูลจริง จึงต้องใช้การตรวจสอบการจำแนกเบื้องต้น โดยการตรวจสอบความสมเหตุสมผล (Logically) ของแบบแผนการตอบที่คาดหวังทั้งหมดกับแบบแผนการตอบข้อสอบที่สังเกตได้ถ้าแบบแผนการตอบข้อสอบที่คาดหวังมีความสมเหตุสมผลกับแบบแผนการตอบข้อสอบที่สังเกตได้แสดงว่าผู้สอบมีคุณลักษณะตามแบบแผนการตอบข้อสอบที่คาดหวัง แต่ถ้าตรวจแล้วพบว่าไม่สมเหตุสมผลกับแบบแผนการตอบข้อสอบที่คาดหวังจะทำการคำนวณความน่าจะเป็น ตัวอย่างเช่น แบบแผนการตอบข้อสอบที่สังเกตได้ของนางสาว ก. เป็น 111111111111101 เมื่อเปรียบเทียบกับแบบแผนการตอบที่คาดหวังจำนวน 16 แบบแผน พบว่ามีความคลาดเคลื่อนที่ไม่สมเหตุสมผลเกิดขึ้น 2 แบบแผน ได้แก่แบบแผนที่ 15 คือ 101101101101101 และแบบแผนที่16 คือ 111111111111111 ที่พบว่า เกิดความคลาดเคลื่อนจาก 1→0 ในข้อที่ 14 จำนวน 1 จุด เมื่อคำนวณความน่าจะเป็นพบว่า แบบแผนที่ 15 มีค่าความน่าจะเป็นเท่ากับ 0.6835 ที่ระดับความสามารถ 1.54 มีคุณลักษณะของผู้สอบเป็น 110111 ส่วนแบบแผนที่ 16 มีค่าความน่าจะเป็นเท่ากับ 0.0318 ที่ระดับความสามารถ 2.37 มีคุณลักษณะของผู้สอบเป็น 111111 ดังนั้นแบบแผนการตอบข้อสอบที่สังเกตได้ของนางสาว ก. จึงมีความเป็นไปได้ที่จะสอดคล้องกับแบบแผนการตอบข้อสอบที่คาดหวังแบบแผนที่ 15 มากกว่าแบบแผนที่ 16 หมายความว่านางสาว ก. มีคุณลักษณะตามแบบแผนการตอบข้อสอบที่คาดหวัง ตั้งแต่แบบแผนที่ 2 ถึงแบบแผนที่ 14 โดยสอดคล้องกับแบบแผนที่ 15 มากที่สุด แต่มีคุณลักษณะไม่ครบตามที่กำหนดในแบบแผนที่กำหนดในแบบแผนที่ 16

3. งานวิจัยที่เกี่ยวกับการนำโมเดลลำดับขั้นคุณลักษณะไปใช้

โมเดลลำดับขั้นคุณลักษณะเป็นโมเดลที่ได้รับการพัฒนาและนำเสนอแนวคิดให้เห็นเป็นรูปธรรม ในปี ค.ศ. 2002 ซึ่งงานวิจัยจะจำกัดอยู่เฉพาะในมหาวิทยาลัยอัลเบอร์ตา (University of Alberta) ซึ่งเป็นที่ทำงานของผู้เสนอแนวคิด จากการศึกษางานวิจัยที่เกี่ยวข้องกับการนำโมเดลลำดับขั้นคุณลักษณะไปใช้ มีดังนี้

3.1 การใช้โมเดลลำดับขั้นคุณลักษณะในการกำหนดมาตรฐานการทดสอบ

Sadesky and Gushta (2004) ได้ประยุกต์โมเดลลำดับขั้นคุณลักษณะในการกำหนดมาตรฐานการสอบวัดผลสัมฤทธิ์วิชาคณิตศาสตร์ ซึ่งเป็นแบบทดสอบย่อยในโปรแกรมบ่งชี้ผลสัมฤทธิ์ของโรงเรียน (School Achievement Indicators Program: SAIP) จัดขึ้นในปี ค.ศ. 1977 เพื่อแสดงให้เห็นว่าโมเดลลำดับขั้นคุณลักษณะเป็นโมเดลที่มีประโยชน์ในการจำแนกความสามารถของนักเรียนในแต่ละระดับมาตรฐาน ได้อย่างมีประสิทธิภาพแบบทดสอบที่ใช้ มีทั้งหมด 125 ข้อ จำแนกระดับของข้อสอบเป็น 5 ระดับในแต่ละระดับมีจำนวนข้อสอบ 25 ข้อ จำแนกระดับ

การปฏิบัติจากโมเดลลำดับขั้นคุณลักษณะโดยใช้โปรแกรม Mathematica 4.1 ที่เขียนโดย Steve Hunka ในแต่ละระดับกำหนดแบบแผนการตอบข้อสอบและจำนวนข้อสอบที่ตอบถูกขั้นต่ำ เช่น ในระดับที 2 ผู้สอบจะตอบข้อสอบถูกโดยมีจำนวนขั้นต่ำในแต่ละเนื้อหาดังนี้ 1) 6 ใน 10 ข้อในเรื่องจำนวนและปฏิบัติการเกี่ยวกับจำนวน (Number and Operation) 2) 1 ใน 3 ข้อ ในเรื่องพีชคณิต (Algebra) 3) 5 ใน 8 ข้อในเรื่องการวัดและเรขาคณิต (Measurement and Geometry) 4) 1 ใน 3 ในเรื่องการจัดการข้อมูลและสถิติ (Data Management and Statistics) และ 5) 15 ใน 25 ข้อจากข้อสอบทั้งหมด

ผลการศึกษาชี้ให้เห็นว่า โมเดลลำดับขั้นคุณลักษณะเป็นเทคนิคที่มีประโยชน์ในการกำหนดมาตรฐาน ดังนี้ 1) แสดงให้เห็นถึงความรู้หรือทักษะที่ผู้สอบมีในแต่ละระดับของการบรรลุมาตรฐานการปฏิบัติ 2) ลดข้อจำกัดในเรื่องความน่าเชื่อถือของคณะผู้ประเมินในการกำหนดค่าความน่าจะเป็นที่ผู้สอบจะตอบข้อสอบถูกในแต่ละระดับ และ 3) ลดข้อจำกัดในเรื่องการใช้คะแนนรวมเอกมิติ (Uni-dimensional Total Score)

3.2 การศึกษาสถานการณ์จำลอง (Simulation Study) เพื่อประเมินวิธีการจำแนกโมเดลลำดับขั้นคุณลักษณะ

Cui Leighton and Zheng (2006) ได้จำลองข้อมูลเพื่อประเมินวิธีการจำแนกที่ใช้ในโมเดลลำดับขั้นคุณลักษณะจำนวน 2 วิธี ได้แก่ วิธี A และวิธี B ลักษณะของลำดับขั้นที่ใช้ในการจำลองข้อมูลมีจำนวน 4 ลักษณะ ได้แก่ ลำดับขั้นเชิงเส้น ลำดับขั้นเชิงลู่เข้า ลำดับขั้นเชิงจำแนก และลำดับขั้นแบบไม่มีโครงสร้าง แต่ละลักษณะมีจำนวน 7 คุณลักษณะ การประเมินจำแนกเป็น 2 ระดับ ได้แก่ ระดับแบบแผนคุณลักษณะ (The Attribute Patterns Level) และระดับคุณลักษณะเป็นรายบุคคล (The Individual Attribute Level) โดยจะแยกเปรียบเทียบแบบแผนจริงสำหรับแต่ละเวกเตอร์การตอบ ผลการวิจัยปรากฏว่า วิธี B ให้ผลการประเมินดีกว่าวิธี A ทั้งระดับแบบแผนคุณลักษณะและคุณลักษณะที่เป็นรายบุคคล และยังชี้ให้เห็นว่าการประเมินระดับคุณลักษณะของบุคคลจากทั้ง 2 วิธี มีความสำคัญมากกว่าระดับแบบแผนคุณลักษณะ

3.3 การประยุกต์ใช้โมเดลลำดับขั้นคุณลักษณะในการประเมินทางวิศวกรรม โดยใช้โมเดลความเกี่ยวข้อง (Connectionist Model) เพื่อประเมินแบบแผนการตอบข้อสอบ

Gierl, Cui and Hunka (2007) ได้ประยุกต์โมเดลลำดับขั้นคุณลักษณะในการประเมินทางวิศวกรรม ซึ่งเป็นนวัตกรรมใหม่ในการวัดเมื่อวิศวกรรมเป็นหลักการที่ใช้เพื่อออกแบบการวิเคราะห์ในการให้คะแนนการรายงานผลการประเมินทางวิศวกรรมประกอบด้วยขั้นตอนสำคัญ 3 ขั้นตอน ขั้นตอนที่ 1 เป็นการสร้างโมเดลการปฏิบัติงานที่มีลักษณะเฉพาะ ขั้นตอนที่ 2 สร้างแผนแม่แบบ (Templates) การปฏิบัติงานตามกรอบของโมเดล และขั้นตอนที่ 3 ใช้โมเดลการวัดทางจิตในการเก็บรวบรวมข้อมูลการตอบของนักเรียนโดยใช้แบบแผนที่สร้างขึ้นในการให้คะแนน การวิจัยครั้งนี้มีเป้าหมายเพื่อบรรยายกระบวนการวิเคราะห์ในขั้นตอนการจำแนกแบบแผนการตอบของผู้สอบโดยใช้ Neural Network ในการคำนวณความน่าจะเป็นของคุณลักษณะ กลุ่มตัวอย่างได้แก่ นักเรียนที่เขียนตอบแบบทดสอบพีชคณิต ในการสอบ Scholastic Aptitude Test (SAT) การดำเนินการวิจัยกำหนดการวิเคราะห์เป็น 2 ลักษณะ ลักษณะแรกกำหนดให้คำนวณเครือข่ายงานโดยไม่มีผลจากภายนอก ลักษณะที่ 2 กำหนดให้คำนวณเครือข่ายงานโดยมีข้อมูลภายนอกซึ่งเป็นการประมาณค่า

ความสามารถสำหรับแต่ละแบบแผนการตอบ ผลการวิจัยปรากฏว่าลักษณะที่ 2 มี Root Mean Square น้อยกว่าลักษณะที่ 1 แสดงให้เห็นว่าข้อมูลจากภายนอกช่วยเพิ่มความสามารถในการอธิบายเครือข่ายของการแก้ปัญหาได้ (Network Solution)

3.4 การประยุกต์โมเดลลำดับขั้นคุณลักษณะและดัชนี Hierarchy Consistency Index (HCI) ในการทดสอบโมเดลทางปัญญาที่ใช้นักเรียนเป็นฐานและผู้เชี่ยวชาญเป็นฐาน

Leighton, Cui and Cor (2008) ได้ประยุกต์โมเดลลำดับขั้นคุณลักษณะและดัชนี Hierarchy Consistency Index (HCI) ในการทดสอบโมเดลทางปัญญาที่ใช้นักเรียนเป็นฐานและผู้เชี่ยวชาญเป็นฐานโดยมีวัตถุประสงค์เพื่อเปรียบเทียบความเหมาะสมของโมเดลทางปัญญาที่สร้างจากนักเรียนและที่สร้างจากผู้เชี่ยวชาญในการทำนายการปฏิบัติงานของนักเรียนจากการทำข้อสอบพีชคณิต 1 และ 2 ในการทดสอบ Scholastic Aptitude Test (SAT) ผลการวิจัยชี้ให้เห็นว่า โมเดลของนักเรียนทำนายการปฏิบัติงานของนักเรียนได้ดีสำหรับกลุ่มที่มีความสามารถปานกลางถึงสูง ส่วนโมเดลของผู้เชี่ยวชาญทำนายการปฏิบัติงานของนักเรียนได้ดีสำหรับกลุ่มที่มีความสามารถปานกลางถึงสูง และกลุ่มที่มีความสามารถปานกลาง หรือเท่ากับค่าเฉลี่ย แสดงให้เห็นว่าโมเดลของผู้เชี่ยวชาญทำนายกระบวนการตอบโดยทั่วไปได้มากกว่าเมื่อผู้สอบมีระดับความสามารถหลากหลาย ส่วนโมเดลของนักเรียนจะทำนายได้ดีกว่าเมื่อผู้สอบมีระดับความสามารถเฉพาะ

3.5 การประยุกต์โมเดลลำดับขั้นคุณลักษณะในการจำแนก และตีความแตกต่างของทักษะทางปัญญาระหว่างกลุ่ม

Gierl, Zheng and Cui (2008) ได้ประยุกต์โมเดลลำดับขั้นคุณลักษณะในการจำแนกและตีความแตกต่างทางปัญญาระหว่างกลุ่ม กลุ่มตัวอย่างเป็นนักเรียนที่เขียนตอบแบบทดสอบพีชคณิต ในการสอบ Scholastic Aptitude Test (SAT) จำนวน 3016 คน เพศหญิงและเพศชายเพศละ 1508 คน ขั้นตอนการทดสอบการทำหน้าที่ต่างกันของระดับคุณลักษณะจำแนกเป็น 4 ขั้นตอน ขั้นตอนที่ 1 ตั้งสมมติฐานเฉพาะของฟังก์ชั่นการทำหน้าที่ต่างกันของระดับคุณลักษณะ (Attribute Level Differential Functioning) ขั้นตอนที่ 2 ประมาณค่าความน่าจะเป็นสำหรับการศึกษาคุณลักษณะ ขั้นตอนที่ 3 ให้คำจำกัดความการจับคู่คุณลักษณะ และขั้นตอนที่ 4 ทดสอบฟังก์ชั่นการทำหน้าที่ต่างกันของระดับคุณลักษณะ กำหนดสมมติฐานการทดสอบจากลำดับขั้นย่อยของแผนภาพลำดับขั้นของคุณลักษณะในการปฏิบัติทางพีชคณิต ซึ่งมีคุณลักษณะที่1 เป็นคุณลักษณะที่ต้องมาก่อนคุณลักษณะที่ 6 และคุณลักษณะที่ 6 เป็นคุณลักษณะที่ต้องมีมาก่อนคุณลักษณะที่ 7 จำแนกสมมติฐานเป็น 2 ข้อ ข้อแรกเป็นการทดสอบความแตกต่างระหว่างกลุ่มสำหรับคุณลักษณะที่ 6 เรื่อง สมการเชิงเส้นตรง (Linear Equation) และข้อที่ 2 เป็นการทดสอบความแตกต่างระหว่างกลุ่มสำหรับคุณลักษณะที่ 7 เรื่อง สมการกำลังสอง (Quadratic Equation) ผลจากการวิจัยปรากฏว่า ไม่มีความแตกต่างระหว่างเพศสำหรับคุณลักษณะที่ 1 และคุณลักษณะที่ 6 แต่มีความแตกต่างระหว่างเพศ สำหรับคุณลักษณะที่ 7 แสดงให้เห็นว่า เพศหญิงและเพศชายมีทักษะพื้นฐานทางคณิตศาสตร์สามารถเทียบเคียงกันได้และสามารถนำไปดำเนินการแก้สมการเชิงเส้นตรงได้เหมือนกัน แต่ต่างกันที่การนำไปแก้สมการกำลังสอง

324870

ตอนที่ 3 การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์

1. การประยุกต์คอมพิวเตอร์ในการทดสอบแบบปรับเหมาะ

การนำคอมพิวเตอร์มาใช้ในการทดสอบ (Computer Assisted Testing) เพื่อเพิ่มประสิทธิภาพในการตรวจข้อสอบ การตัดเกรด การรายงานผลการสอบ และการสร้างคลังข้อสอบ มีการดำเนินการอย่างแพร่หลายระหว่างปี ค.ศ.1970 ถึงปี ค.ศ.1980 สำหรับการทดสอบแบบปรับเหมาะเริ่มมีการประยุกต์คอมพิวเตอร์มาใช้ในโครงการวิจัยการทดสอบแบบปรับเหมาะโดยได้รับการสนับสนุนจาก U.S. Office of Naval Research ตั้งแต่ปลายปี ค.ศ.1960 ถึงต้นปี ค.ศ. 1970 (Weiss, 1983)

คอมพิวเตอร์เป็นเทคโนโลยีที่มีหน่วยความจำขนาดใหญ่สามารถเก็บรวบรวมข้อมูลทั้งตัวหนังสือและรูปภาพได้จำนวนมาก สามารถประมวลผลได้อย่างรวดเร็วรวมทั้งมีการปฏิสัมพันธ์กับผู้สอบได้อย่างมีประสิทธิภาพ คอมพิวเตอร์จึงมีความเหมาะสมและมีประโยชน์อย่างยิ่งต่อการนำมาใช้ในการทดสอบแบบปรับเหมาะที่ต้องมีระบบการส่งผ่าน ซึ่งสามารถนำข้อสอบออกมาจากคลังข้อสอบ และนำให้ผู้สอบตอบได้อย่างรวดเร็ว โดยต้องมีหลักเกณฑ์ในการคัดเลือกข้อสอบเริ่มต้น การประมาณค่าความสามารถของผู้สอบเพื่อคัดเลือกข้อสอบระหว่างกระบวนการทดสอบและยุติการทดสอบที่ชัดเจน (Van der Linden, 1999)

การประยุกต์คอมพิวเตอร์ในการทดสอบแบบปรับเหมาะ เป็นการดำเนินการสอบตามระดับความสามารถของผู้สอบด้วยคอมพิวเตอร์ ซึ่งสามารถจัดทำได้ในทุกกระบวนของการทดสอบ ตั้งแต่การสร้างคลังข้อสอบ การวิเคราะห์สารสนเทศของข้อสอบ การคัดเลือกข้อสอบ การประมาณค่าความสามารถของผู้สอบ จนกระทั่งยุติการทดสอบ ข้อสอบจะถูกเก็บไว้ในคอมพิวเตอร์และแสดงออกทางจอภาพ (Monitor) เพื่อให้ผู้สอบตอบข้อสอบผ่านทางแป้นพิมพ์ (Keyboard) หรือเม้าส์ (Mouse) โปรแกรมคอมพิวเตอร์จะทำหน้าที่เสมือนนักจิตวิทยาที่ดำเนินการสอบโดยใช้แบบทดสอบเชาว์ปัญญาของบิเนท์ซึ่งจะช่วยกำหนดข้อสอบเริ่มต้น ข้อสอบที่เหมาะสมกับความสามารถของผู้สอบในระหว่างกระบวนการทดสอบ และตัดสินยุติการทดสอบ (Weiss, 2004)

สำหรับการหาค่าสารสนเทศ (Parameter) ของข้อสอบที่จะนำมาใช้กับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ (Computerized Adaptive Testing: CAT) ภายใต้โมเดลทฤษฎีการตอบสนองข้อสอบ (Item Response Theory Model) ส่วนใหญ่จะใช้การหาค่าสารสนเทศของข้อสอบโดยใช้ Graded Response Model (GRM) ต่อมาในปี ค.ศ. 2011 Smits, Cuijpers and Stranten ได้นำหลักการของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์มาประยุกต์กับงานวิจัยทางด้านจิตวิทยา ซึ่งมีการหาค่าพารามิเตอร์ของข้อสอบโดยใช้การวิเคราะห์องค์ประกอบเชิงยืนยัน (Confirmatory Factor Analysis: CFA) โดยมีการกำหนดค่า Comparative Fit Index (CFI) มากกว่า 0.95 ค่า Tucker-Lewis Index (TLI) มากกว่า 0.95 ค่า Standardized Root Mean Square Residual (SRMR) น้อยกว่า 0.08 และค่าเฉลี่ยความสัมพันธ์น้อยกว่า 0.10 ร่วมกับ Graded Response Model (GRM) เพื่อให้มีความเหมาะสม และสอดคล้องกับหลักการทางด้านจิตวิทยา

2. องค์ประกอบของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์

การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ (Computerized Adaptive Testing: CAT) มีองค์ประกอบสำคัญ ดังนี้

2.1 การสร้างคลังข้อสอบ (Item Bank Creating)

การสร้างคลังข้อสอบจะต้องคำนึงถึงมาตรฐานของข้อสอบด้วย เพื่อให้ได้คลังข้อสอบที่มีข้อสอบที่มีประสิทธิภาพและมีมาตรฐาน ซึ่งมาตรฐานในการสร้างข้อสอบหรือแบบทดสอบนั้นมี 5 มาตรฐาน (สถาบันการทดสอบแห่งชาติ, 2555) ดังนี้

1) กำหนดจุดมุ่งหมาย เนื้อหาสาระ และขอบเขตของการทดสอบให้ชัดเจน

2) กำหนดรูปแบบ หรือลักษณะเฉพาะที่มุ่งวัด (Construct) ให้ชัดเจน เช่น องค์ประกอบที่ต้องการวัด การวิเคราะห์ขอบเขตของความรู้ที่ต้องการวัด เป็นต้น

3) จัดทำเอกสารแสดงกระบวนการสร้างข้อสอบ หรือแบบทดสอบ เพื่อให้การสร้างข้อสอบหรือแบบทดสอบมีคุณภาพตามข้อกำหนดของคุณลักษณะของข้อสอบ

4) ตรวจสอบความถูกต้องของข้อสอบให้ตรงตามคุณลักษณะเฉพาะ และครอบคลุมเนื้อหาที่กำหนด

5) ตรวจสอบคุณภาพของข้อสอบ หรือแบบทดสอบตามหลักวิชาการของมาตรฐานความยุติธรรมและคุณภาพ โดยมีผู้เชี่ยวชาญในเนื้อหาที่กำหนดเป็นผู้ตรวจสอบ

การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ จะต้องสร้างคลังข้อสอบที่มุ่งวัดคุณลักษณะหรือความสามารถที่ต้องการได้ ครอบคลุมระดับความสามารถที่ต่าง ๆ กัน และมีข้อสอบในคลังข้อสอบอย่างน้อย 100 – 200 ข้อ (Weiss, 1988) ซึ่งคลังข้อสอบที่ดีจะต้องมีจำนวนข้อสอบเท่ากับ 2^n เมื่อ n คือ จำนวนข้อสอบที่ต้องการใช้ทดสอบ เช่น ถ้าต้องการสอบประมาณ 10 ข้อ ต้องมีข้อสอบในคลังข้อสอบ 1,024 ข้อ เป็นต้น หากต้องการสอบแบบกำหนดจำนวนข้อที่จะใช้คงที่เหมือนกันทุกคน จำนวนข้อสอบในคลังข้อสอบควรมีข้อสอบอย่างน้อย $n(n+1) / 2$ ข้อ เช่น ถ้าต้องการสอบ 10 ข้อ ต้องมีข้อสอบในคลังข้อสอบ 55 ข้อ เป็นต้น (Allen & Yen, 1979) โดยข้อสอบในแต่ละข้อควรมีค่าอำนาจจำแนกระหว่าง 0.80 ถึง 2.50 มีค่าความยากที่ครอบคลุมระดับช่วงต่าง ๆ อย่างเหมาะสม มีค่าระหว่าง -2.50 ถึง +2.50 ค่าการเดาไม่ควรเกิน 0.30 และควรมีข้อสอบในจำนวนที่เท่า ๆ กัน ในแต่ละระดับความยาก (Urry, 1977)

Molina, Pareja and Sanmartin (2008) ได้เสนอขั้นตอนการสร้างคลังข้อสอบที่จะนำมาใช้ในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ เพื่อให้มีประสิทธิภาพมากขึ้น ประกอบด้วย 5 ขั้นตอน ดังนี้

1) กำหนดกฎเกณฑ์ คำอธิบายต่าง ๆ รวมถึงคำสั่งที่เกี่ยวข้องของข้อสอบ ที่คาดว่าจะส่งผลทางจิตวิทยาระหว่างดำเนินการสอบ

2) คัดเลือกข้อสอบที่เหมาะสม และตรงตามกฎเกณฑ์ที่กำหนดในคลังข้อสอบที่ได้จากการบริหารจัดการสอบโดยใช้คอมพิวเตอร์ หรือการจัดการสอบโดยใช้กระดาษคำตอบ

3) พัฒนาในฐานข้อมูลของคลังข้อสอบ จากชุดข้อสอบที่คัดเลือกจากการจัดการสอบโดยใช้คอมพิวเตอร์ หรือการจัดการสอบโดยใช้กระดาษคำตอบ เพื่อนำมาใช้ในการบริหารการสอบแบบปรับเหมาะด้วยคอมพิวเตอร์

4) แยกข้อสอบที่ได้จากจากการบริหารจัดการสอบโดยใช้คอมพิวเตอร์ หรือการจัดการสอบโดยใช้กระดาษคำตอบ และข้อสอบที่คัดเลือกแล้วไปจัดเก็บคนละส่วนกัน และนำข้อสอบที่ทำการคัดเลือกมาตัดสินใจคัดเลือกอีกครั้ง จึงนำไปพัฒนาเป็นคลังข้อสอบที่จะนำไปใช้

5) วิเคราะห์ความแตกต่างของผลกระทบทางจิตวิทยาของกลุ่มข้อสอบที่ได้จากการบริหารจัดการสอบโดยใช้คอมพิวเตอร์ หรือการจัดการสอบโดยใช้กระดาษคำตอบกับกลุ่มข้อสอบในคลังข้อสอบที่พัฒนาขึ้น ซึ่งสามารถนำไปเป็นข้อมูลในการพัฒนาปรับปรุงคลังข้อสอบใหม่ให้มีคุณภาพมากยิ่งขึ้น

2.2 การคัดเลือกข้อสอบข้อเริ่มต้น (First Item Selection)

Wainer (1990) กล่าวว่า การทดสอบควรเริ่มต้นจากการเลือกข้อสอบที่มีระดับความยากที่เท่ากับหรือต่ำกว่าระดับความสามารถของผู้สอบ เพื่อให้การประมาณค่าความสามารถของผู้สอบได้ถูกต้องมากขึ้น นอกจากนั้น การเลือกข้อสอบข้อเริ่มต้นอาจจะพิจารณาผลการเรียนของผู้สอบที่ผ่านมาก็ได้ สำหรับการแบ่งวิธีการเลือกข้อสอบได้ 2 กรณี ดังนี้

2.2.1 ประชากรผู้สอบมีความสามารถที่ใกล้เคียงกัน หรือเป็นประชากรของผู้สอบที่ไม่มีข้อมูลสารสนเทศระดับการศึกษา ควรเลือกข้อสอบที่มีระดับความยากปานกลางเป็นข้อสอบข้อเริ่มต้น

2.2.2 ประชากรผู้สอบมีความสามารถค่อนข้างแตกต่างกัน หรือเป็นประชากรผู้สอบที่พอมีข้อมูลสารสนเทศระดับการศึกษาของผู้สอบ ควรเริ่มต้นที่ข้อสอบที่มีความยากระดับปานกลางในแต่ละระดับการศึกษาของผู้สอบ

2.3 การคัดเลือกข้อสอบข้อถัดไป (Next Item Selection)

การคัดเลือกข้อสอบข้อถัดไปของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์เป็นขั้นตอนที่มีความสำคัญขั้นตอนหนึ่ง เนื่องจากข้อสอบที่เลือกมาเป็นข้อสอบข้อถัดไปนั้นส่งผลต่อการประมาณค่าความสามารถของผู้สอบ ดังนั้น ควรเลือกข้อสอบที่มีความสอดคล้องกับความสามารถของผู้สอบขณะทำการทดสอบ ซึ่งวิธีการคัดเลือกข้อสอบข้อถัดไปที่นิยมใช้ คือ วิธีการคัดเลือกข้อสอบข้อถัดไป โดยการพิจารณาจากค่าความยากของข้อสอบ (b – Parameter) ซึ่งจะคัดเลือกข้อสอบที่มีค่าความยากของข้อสอบใกล้เคียงกับค่าความสามารถของผู้สอบขณะทำการทดสอบมาเป็นข้อสอบถัดไป และวิธีการคัดเลือกข้อสอบข้อถัดไป โดยพิจารณาตำแหน่งสารสนเทศสูงสุดของข้อสอบ ซึ่งจะคัดเลือกข้อสอบที่มีค่าสารสนเทศของข้อสอบใกล้เคียงกับค่าความสามารถของผู้สอบมาเป็นข้อสอบถัดไป สามารถเรียกอีกอย่างหนึ่งว่า การคัดเลือกข้อสอบข้อถัดไป โดยใช้ตำแหน่งสารสนเทศสูงสุดที่ใกล้เคียงกับค่าความสามารถของผู้สอบ (Maximum Information: MI)

การคัดเลือกข้อสอบข้อถัดไปของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ควรเลือกข้อสอบที่มีความสอดคล้องกับวิธีการประมาณค่าความสามารถของผู้สอบ (Hulin, Drasgow, & Parson, 1983) มีนักวิชาการนำเสนอวิธีการคัดเลือกข้อสอบไว้ ดังนี้

2.3.1 การคัดเลือกข้อสอบข้อถัดไปที่ระดับความยากของข้อสอบ (b) กล่าวคือ เมื่อประมาณค่าความสามารถของผู้สอบแล้ว จะนำค่าความสามารถของผู้สอบมาเปรียบเทียบกับค่าระดับความยากของข้อสอบทุกข้อ และเลือกข้อสอบที่มีระดับความยากที่ใกล้เคียงกับค่าความสามารถของผู้สอบในขณะนั้นมากที่สุดมาเป็นข้อสอบข้อถัดไป โดยไม่คำนึงถึงค่าอำนาจจำแนกของข้อสอบ (a) และค่าการเดาข้อสอบ (c)

2.3.2 การคัดเลือกข้อสอบข้อถัดไป โดยใช้ตำแหน่งสารสนเทศสูงสุดที่ใกล้เคียงกับค่าความสามารถของผู้สอบ (Maximum Information) กล่าวคือ การคัดเลือกข้อสอบด้วยวิธีนี้จะมี

การพิจารณาปัจจัยของค่าโอกาสการเดาข้อสอบร่วมด้วย ซึ่งจะคัดเลือกข้อสอบที่มีค่าสารสนเทศสูงสุดที่ใกล้เคียงกับค่าประมาณความสามารถของผู้สอบ มาเป็นข้อสอบข้อถัดไป ค่าสารสนเทศสูงสุดของข้อสอบสำหรับโมเดล 3 – Parameters สามารถคำนวณได้จากสมการ

$$m_i = b_i + \frac{1}{D_{a_i}} In \left[\frac{1 + \sqrt{1 + 8c_i}}{2} \right]$$

เมื่อ m_i คือ ค่า θ ณ ตำแหน่งที่ข้อสอบข้อนั้นให้สารสนเทศสูงสุด
D คือ ค่าคงที่
ln คือ natural logarithm
a_i คือ ค่าอำนาจจำแนกของข้อสอบ
b_i คือ ค่าความยากของข้อสอบ
c_i คือ ค่าโอกาสการเดาข้อสอบถูก

2.4 การประมาณค่าความสามารถ (Ability Estimation)

การประมาณค่าความสามารถของผู้สอบสำหรับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์มีวิธีคำนวณ ดังนี้

2.4.1 ประมาณค่าความสามารถของผู้สอบด้วยวิธีของเบส์

การประมาณค่าความสามารถของผู้สอบด้วยวิธีของเบส์ (Bayesian) มีการประมาณค่าความสามารถของผู้สอบได้จาก Bayes's theorem ซึ่งสามารถคำนวณได้ จากสมการ

$$f(\theta/U) = K.L(U/\theta).f(\theta)$$

เมื่อ $f(\theta/U)$ คือ Posterior Distribution ของ θ
$L(U/\theta)$ คือ Likelihood Function ของ Vector U (Item Response)
$f(\theta)$ คือ Prior Distribution ของ θ
K คือ Constant

2.4.2 การประมาณค่าความสามารถของผู้สอบด้วยวิธีของเบส์ปรับใหม่

การประมาณค่าความสามารถของผู้สอบด้วยวิธีของเบส์ปรับใหม่ (Bayesian Updating) ทำให้การประมาณค่าความสามารถของผู้สอบได้ค่าค่อนข้างคงที่ (Owen, 1975) ซึ่งสามารถคำนวณได้ ดังนี้

1) กรณีที่ตอบข้อสอบถูก

มหาวิทยาลัยบูรพา Burapha University

$$\theta_{m+1} = \theta_m + (1-c)\left[\frac{\sigma_m^2}{\sqrt{\frac{1}{a^2}+\sigma_m^2}}\right]\left[\frac{O(D)}{c+(1-c)\Lambda(-D)}\right]$$

$$\sigma_{m+1}^2 = \sigma_m^2\left[1-\left[\frac{1-c}{1+\frac{1}{a^2\sigma_m^2}}\right]\left[\frac{O(D)}{B}\right]\left[\frac{(1-c)O(D)}{B}-D\right]\right]$$

เมื่อ $$D = \frac{b-\theta_m}{\sqrt{\frac{1}{a^2}+\sigma^2}}$$

และ $$B = c + (1-c)\times\Lambda(-D)$$

2) กรณีที่ตอบข้อสอบผิด

$$\theta_{m+1} = \theta_m - \left[\frac{\sigma_m^2}{\sqrt{\frac{1}{a^2}+\sigma_m^2}}\right]\left[\frac{O(D)}{A(D)}\right]$$

$$\sigma_{m+1}^2 = \sigma_m^2\left[1-\left[\frac{O(D)}{1+\frac{1}{a^2\sigma_m^2}}\right]\frac{\left[\frac{O(D)}{A(D)}+D\right]}{A(D)}\right]$$

เมื่อ

θ_m คือ ความสามารถของผู้สอบที่ประมาณได้ก่อนตอบข้อสอบข้อ m+1 โดยกำหนดค่า θ_m เท่ากับ 0.000 เมื่อเริ่มต้นการทดสอบ

σ_m^2 คือ ค่าความแปรปรวนในการประมาณค่าความสามารถของผู้สอบก่อนตอบข้อสอบข้อ m+1 โดยกำหนดค่า σ_m^2 เท่ากับ 1.000 เมื่อเริ่มต้นการทดสอบ

θ_{m+1} คือ ความสามารถของผู้สอบที่ประมาณได้หลังตอบข้อสอบข้อ m+1 แล้ว

σ_{m+1}^2 คือ ค่าความแปรปรวนในการประมาณค่าความสามารถของผู้สอบหลังตอบข้อสอบข้อ m+1

a_i คือ ค่าอำนาจจำแนกของข้อสอบข้อที่ m+1

b_i คือ ค่าความยากของข้อสอบข้อที่ m+1

มหาวิทยาลัยบูรพา Burapha University

c_i คือ ค่าโอกาสการเดาของข้อสอบข้อที่ m+1

D คือ จุดตัดบนแกน X

O(D) คือ ค่า Ordinate ของโค้งปกติที่จุด D

A(D) คือ พื้นที่ใต้โค้งปกติจากค่า D ต่ำสุดจนถึงจุด D

2.4.3 การประมาณค่าความสามารถของผู้สอบด้วยวิธีความน่าจะเป็น

การประมาณค่าความสามารถของผู้สอบด้วยวิธีความน่าจะเป็น (Maximum Likelihood) มีอยู่หลายวิธี แต่วิธีที่นิยมมากที่สุด คือ วิธีความน่าจะเป็นแบบมีเงื่อนไข (Conditional Maximum Likelihood) โดยการประมาณค่าความสามารถของผู้สอบโดยวิธีนี้ยังมีข้อจำกัดอยู่ คือ ถ้าผู้สอบตอบข้อสอบถูกทุกข้อ หรือผิดทุกข้อ จะไม่สามารถประมาณค่าความสามารถของผู้สอบได้ (Hambleton & Swaminathan, 1985) วิธีการประมาณค่าความสามารถของผู้สอบสำหรับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ของวิธีความน่าจะเป็นแบบมีเงื่อนไข มีขั้นตอน ดังนี้

1) ประมาณค่าความสามารถเริ่มต้น ($\theta_m = 0; m = 0$)

$$\theta_0 = \ln\left(\frac{r_a}{k - r_a}\right)$$

เมื่อ

$r_a = \sum a_i u_i$

$u_i = 1$ เมื่อตอบข้อสอบถูก

$u_i = 0$ เมื่อตอบข้อสอบผิด

$a_i =$ ค่าอำนาจจำแนกของข้อสอบข้อที่ i

$k =$ จำนวนข้อสอบทั้งหมดที่มีผู้สอบตอบ

2) หาค่า $P_i(\theta_m)$ และ $Q_i(\theta_m)$

$$P_i(\theta_m) = c_i + (1 - c_i)\frac{e^{Da_i(\theta_m - b_i)}}{Da_i(\theta_m - b_i)}$$

เมื่อ

$\theta_m =$ ความสามารถของผู้สอบที่ประมาณค่าได้ในครั้งที่ m

$a_i =$ ค่าอำนาจจำแนกของข้อสอบข้อที่ i

$b_i =$ ค่าความยากของข้อสอบข้อที่ i

$c_i =$ ค่าโอกาสในการเดาข้อสอบข้อที่ i

$D =$ ค่าคงที่ มีค่าประมาณ 1.7

$e =$ ค่าคงที่ มีค่าประมาณ 2.7182818

3) หาค่าปรับแก้ h_m

มหาวิทยาลัยบูรพา Burapha University

$$h_m = \frac{D[r_a - \sum P_i(\theta_m)]}{-D^2 \sum P_i(\theta_m) Q(\theta_m)}$$

4) ประมาณค่าความสามารถของผู้สอบใหม่ θ_{m+1}

$$\theta_{m+1} = \theta_m - h_m$$

5) ประมาณค่า θ ซ้ำจนบรรลุเงื่อนไข โดยที่คำนวณซ้ำขั้นตอนที่ 2 ถึงขั้นตอนที่ 4 จนกระทั่ง h_m เข้าใกล้ศูนย์ ($h_m < 0.001$) หรือบรรลุตามเงื่อนไขจึงยุติการทดสอบ

2.5 เกณฑ์การยุติการทดสอบ (Termination Criteria)

การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์จะสิ้นสุดการทดสอบ ก็ต่อเมื่อผลการทดสอบเป็นไปตามเงื่อนไข หรือเกณฑ์ที่กำหนด เช่น กำหนดจำนวนข้อสอบไว้ล่วงหน้า เวลาที่ใช้ในการทดสอบ คะแนนจุดตัดที่แยกระดับความสามารถของผู้สอบ ความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถของผู้สอบ เป็นต้น สำหรับเกณฑ์การยุติการทดสอบที่นิยมใช้กันมี 2 ประเภท ดังนี้

2.5.1 กำหนดจำนวนข้อสอบคงที่

การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ที่ใช้เกณฑ์การยุติการทดสอบประเภทนี้ ผู้สอบจะต้องทำแบบทดสอบจนครบตามจำนวนข้อสอบที่กำหนด การทดสอบจึงยุติลง เช่น กำหนดให้ผู้สอบทำแบบทดสอบจำนวน 20 ข้อ ผู้สอบต้องทำข้อสอบให้ครบ 20 ข้อตามที่กำหนด เป็นต้น การทดสอบโดยใช้เกณฑ์การยุติการทดสอบประเภทนี้ อาจจะทำให้คุณภาพของการวัดผลมีความแตกต่างกันได้

2.5.2 กำหนดระดับความคลาดเคลื่อนมาตรฐานที่เหมาะสม

การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ที่ใช้เกณฑ์การยุติการทดสอบประเภทนี้ ผู้สอบจะต้องทำแบบทดสอบไปเรื่อย ๆ จนกว่าค่าความคลาดเคลื่อนมาตรฐาน มีค่าลดลงถึงระดับที่กำหนด การทดสอบจึงยุติลง การทดสอบที่ใช้เกณฑ์ประเภทนี้ ผู้สอบแต่ละคนอาจจะทำข้อสอบในจำนวนที่ไม่เท่ากัน หรือข้อสอบที่ไม่เหมือนกันได้ โดยทั่วไปจะกำหนดค่าความคลาดเคลื่อนมาตรฐานน้อยกว่า 0.30 สำหรับการหาค่าความคลาดเคลื่อนมาตรฐาน สามารถคำนวณได้จากสมการ

$$SE(\theta) = \frac{1}{\sqrt{I(\theta)}}$$

เมื่อ

$SE(\theta)$ คือ ความคลาดเคลื่อนมาตรฐานของการประมาณค่า θ

$I(\theta)$ คือ สารสนเทศของแบบทดสอบที่ให้สำหรับผู้มีความสามารถ θ

นอกจากนี้ ยังมีนักวิชาการได้กล่าวถึง องค์ประกอบทั่วไปของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ไว้ ดังนี้

1. จุดเริ่มต้นการทดสอบ (Starting Point) การทดสอบควรเริ่มต้นด้วยข้อสอบที่ตรงกับความสามารถของผู้สอบหรือใกล้เคียงกับความสามารถของผู้สอบมากที่สุด ซึ่งจะช่วยให้จำนวนข้อสอบที่ใช้ในการทดสอบสำหรับแต่ละบุคคลมีจำนวนลดลง (Latu & Chapman, 2002) โดยทั่วไปนิยมเลือกข้อสอบที่มีความยากปานกลาง โดยอาจดำเนินการได้ใน 2 กรณี กรณีแรก กลุ่มผู้สอบมีความสามารถใกล้เคียงกัน (Homogeneous) หรือไม่มีข้อมูลสำหรับผลสัมฤทธิ์ที่ผ่านมาควรเริ่มต้นด้วยข้อสอบที่มีค่าความยากปานกลางเท่ากันสำหรับทุกคน กรณีที่สอง ผู้สอบมีความสามารถแตกต่างกัน (Heterogeneous) หรือมีข้อมูลเกี่ยวกับการศึกษาของผู้สอบบ้าง ควรใช้ค่าความยากปานกลางที่เหมาะสมสำหรับผู้สอบแต่ละกลุ่มความสามารถ โดยพิจารณาจากข้อมูลที่มีอยู่ เช่น กลุ่มอายุ ระดับชั้นที่ศึกษา เป็นต้น (ศิริชัย กาญจนวาสี, 2550)

ความสำคัญของจุดเริ่มต้นยังคงเป็นที่โต้แย้งกันในกลุ่มนักวิจัย บางคนกล่าวว่าจุดเริ่มต้นไม่มีผลกระทบต่อการสอบถ้ามีจำนวนข้อสอบมากพอ เช่น Lord (1977) ได้ทดลองใช้สถานการณ์จำลอง โดยเริ่มต้นด้วยข้อสอบที่มีค่าความยากระดับต่าง ๆ พบว่าเกือบไม่มีผลต่อความถูกต้องแม่นยำในการประมาณค่าความสามารถของผู้สอบเลย โดยเฉพาะกรณีที่มีข้อสอบจำนวนมาก แต่บางคนกล่าวถึงจุดเริ่มต้นที่ไม่เหมาะสม โดยเฉพาะกรณีที่มีข้อสอบในคลังข้อสอบจำนวนมาก แต่บางคนกล่าวว่าจุดเริ่มต้นที่ไม่เหมาะสมจะทำให้ผู้สอบเกิดความวิตกกังวลและรู้สึกผิดหวังซึ่งมีผลกระทบต่อการทำแบบทดสอบ (Latu & Chapman, 2002)

2. การคัดเลือกข้อสอบและการประมาณค่าความสามารถ (Item Selection and Ability Estimate) การคัดเลือกข้อสอบและการประมาณค่าความสามารถของผู้สอบมีประสิทธิภาพมากขึ้นเมื่อนำทฤษฎีการตอบสนองข้อสอบ (Item Response Theory) มาประยุกต์ใช้ โดยทั่วไปนิยมใช้วิธีที่สอดคล้องกัน เพื่อความสะดวกรวดเร็วในการคำนวณ วิธีที่นิยมนำมาใช้ในการประมาณค่าความสามารถมี 2 วิธี ได้แก่ วิธีการประมาณค่าความเป็นไปได้สูงสุด (Maximum Likelihood Estimate) และประมาณค่าแบบเบส์ (Bayestian Estimate) วิธีการคัดเลือกข้อสอบที่ใช้สำหรับการประมาณค่าความเป็นไปได้สูงสุด ประกอบด้วยการคัดเลือกข้อสอบที่มีค่าความยากสอดคล้องกับค่าความสามารถที่ประมาณได้ การคัดเลือกข้อสอบที่มีตำแหน่งสารสนเทศของค่าความยาก ค่าอำนาจจำแนก และค่าการเดาสูงสุด สอดคล้องกับความสามารถที่ประมาณได้และการคัดเลือกข้อสอบที่ให้สารสนเทศสูงสุดที่ตำแหน่งความสามารถนั้น ส่วนวิธีการคัดเลือกข้อสอบที่ใช้สำหรับการประมาณค่าแบบเบส์ใช้การคัดเลือกข้อสอบข้อที่ยังไม่ได้ใช้นำมาใช้ในการทดสอบนั้นจะให้ค่าความแปรปรวนของค่าความสามารถที่คาดหวังมีค่าต่ำสุด (ศิริชัย กาญจนวาสี, 2550)

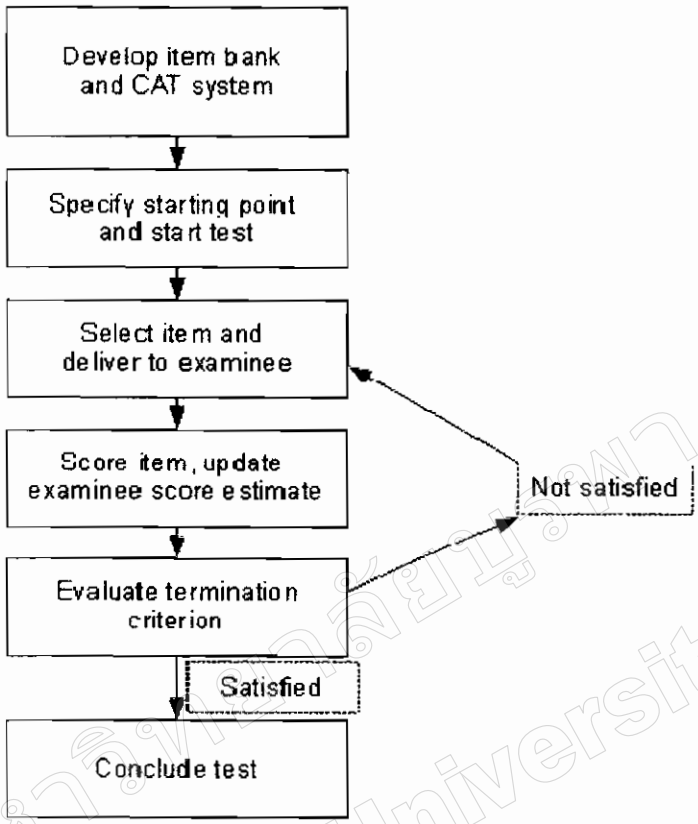
3. เกณฑ์การยุติการทดสอบ (Termination Criteria) ปัจจุบัน มีการพัฒนาวิธีที่ใช้ในการยุติการทดสอบหลายวิธี แต่ละวิธีที่นิยมใช้มี 2 วิธี วิธีแรกคือ การทดสอบจนกระทั่งได้ระดับความคงที่ในการวัด (Level of Measurement Stability or Consistency) โดยการกำหนดระดับความคลาดเคลื่อนที่ยอมรับได้ วิธีที่สองคือ การกำหนดจำนวนข้อสอบให้คงที่วิธีการกำหนดความยาวของแบบสอบให้คงที่ช่วยให้แน่ใจได้ว่าผู้สอบได้ทำข้อสอบในจำนวนข้อสอบที่เท่ากันและช่วยให้แบบทดสอบ

ไม่ยาวเกินแต่ให้คุณภาพของการวัดผลที่แตกต่างกันเนื่องจากมีระดับความยากของข้อสอบแตกต่างกัน การใช้การยุติการทดสอบโดยพิจารณาจากค่าความคลาดเคลื่อนในการวัด
จึงเป็นวิธีที่ดีกว่า (Latu & Chapman, 2002)

4. ขนาดและเนื้อหาของข้อสอบในคลังข้อสอบ (Size and Content of Item Banks) ความตรงของแบบทดสอบขึ้นอยู่กับความเกี่ยวข้องและความเหมาะสมของข้อสอบที่ใช้ แต่สำหรับการทดสอบปรับเหมาะโดยใช้คอมพิวเตอร์ความตรงของแบบทดสอบจะขึ้นอยู่กับทั้งจำนวนและเนื้อหาของข้อสอบในคลังข้อสอบ นอกจากข้อสอบจะต้องมีจำนวนมากเพียงพอในการประมาณค่าระดับความสามารถของผู้สอบซึ่งมีความแตกต่างหลากหลายกันได้อย่างถูกต้องเหมาะสมแล้ว ข้อควรคำนึงที่สำคัญประการหนึ่งก็คือ เนื้อหาของข้อสอบเนื่องจากการมีข้อสอบในคลังจำนวนมากทำให้มีโอกาสที่จะมีข้อสอบที่ไม่ดีได้มาก ซึ่งข้อสอบเหล่านั้นจะมีผลกระทบอย่างมากต่อการให้คะแนนสุดท้ายในการทดสอบแบบปรับเหมาะ เพราะในการทดสอบแบบปรับเหมาะมีการตรวจให้คะแนนในทันที ข้อสอบที่ถูกเลือกให้ผู้สอบทำจึงมีผลต่อความถูกต้องในการประมาณค่าความสามารถของผู้สอบและการให้คะแนนสุดท้ายโดยไม่มีการแก้ไขหรือย้อนกลับได้ (Latu & Chapman, 2002)

3. หลักการของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์

การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์มีหลักการคัดเลือกข้อสอบที่เหมาะสมกับแต่ละบุคคล ซึ่งอยู่บนฐานการตอบข้อสอบข้อที่ผ่านมาของผู้สอบ กล่าวคือ เมื่อผู้สอบทำข้อสอบข้อเริ่มต้น หรือชุดเริ่มต้น ทั้งนี้ขึ้นอยู่กับการออกแบบ แล้วจะนำผลการตอบข้อสอบมาวิเคราะห์ หรือประเมินระดับความสามารถของผู้สอบ เพื่อที่จะเลือกข้อสอบที่เหมาะสมข้อถัดไป โดยอาศัยหลักการคัดเลือกข้อสอบว่า ถ้าทำข้อสอบข้อที่ผ่านมาถูกต้อง ข้อสอบข้อถัดไปจะยากมากขึ้นหากทำข้อสอบข้อที่ผ่านมาผิด ข้อสอบข้อถัดไปจะง่ายลง กระบวนการนี้จะดำเนินการไปจนกว่าจะประเมินระดับความสามารถของผู้สอบเป็นไปตามเกณฑ์การยุติการทดสอบที่กำหนด (ขึ้นอยู่กับการออกแบบ) การทดสอบจึงยุติลง (Green, 1984) โดยขั้นตอนการดำเนินการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ แสดงดังภาพที่ 4



ภาพที่ 4 ขั้นตอนการดำเนินการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ (Thompson, 2011, p. 2)

จากภาพที่ 4 แสดงขั้นตอนการดำเนินการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ซึ่งมี 5 ขั้นตอนสามารถอธิบายได้ ดังนี้

ขั้นตอนที่ 1 การพัฒนาคลังข้อสอบและระบบการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ (Develop Item Bank and CAT System) เป็นขั้นตอนที่ต้องพิจารณาขนาดของคลังข้อสอบและเงื่อนไขต่าง ๆ ที่สอดคล้องกับหลักการของทฤษฎีการตอบสนองข้อสอบ เพื่อคัดเลือกข้อสอบที่ตรงตามเงื่อนไขที่กำหนดมาจัดเก็บในคลังข้อสอบ ซึ่งคลังข้อสอบที่ดีจะต้องมีจำนวนข้อสอบเท่ากับ 2^n เมื่อ n คือ จำนวนข้อสอบที่ต้องการใช้ทดสอบ เช่น ถ้าต้องการทดสอบประมาณ 10 ข้อ ต้องมีข้อสอบในคลังข้อสอบ 1,024 ข้อ เป็นต้น หากต้องการทดสอบแบบกำหนดจำนวนข้อที่จะใช้คงที่เหมือนกันทุกคน จำนวนข้อสอบในคลังข้อสอบควรมีข้อสอบอย่างน้อย $n(n+1) / 2$ ข้อ เช่น ถ้าต้องการทดสอบ 10 ข้อ ต้องมีข้อสอบในคลังข้อสอบ 55 ข้อ เป็นต้น

ขั้นตอนที่ 2 การคัดเลือกข้อสอบข้อเริ่มต้น (Specify Starting Point and Start Test) เป็นการคัดเลือกข้อสอบข้อเริ่มต้น ซึ่งเป็นข้อสอบที่มีระดับความยากปานกลาง กล่าวคือ การเลือกข้อสอบที่มีระดับความยากใกล้เคียงกับค่าเฉลี่ยของค่าความยากของข้อสอบทั้งหมด หรือการคัดเลือกข้อสอบข้อเริ่มต้นจากการสุ่มในช่วงระดับความยากของข้อสอบที่กำหนด หรือการคัดเลือกข้อสอบข้อเริ่มต้นตามกฎเกณฑ์ที่กำหนด

ขั้นตอนที่ 3 การคัดเลือกข้อสอบข้อถัดไป (Select Item and Deliver to Examinee) เป็นขั้นตอนที่ต้องพิจารณาคัดเลือกข้อสอบตามเงื่อนไขที่กำหนด โดยอาศัยผลการตอบข้อสอบข้อก่อนหน้ามาร่วมพิจารณาด้วย

ขั้นตอนที่ 4 การประมาณค่าความสามารถของผู้สอบ (Score Item, Update Examinee Score Estimate) เป็นการประมาณค่าความสามารถของผู้สอบหลังจากตอบข้อสอบแล้ว ซึ่งวิธีการประมาณค่าความสามารถของผู้สอบที่นิยมใช้กัน คือ 1) การประมาณค่าความสามารถของผู้สอบด้วยวิธีของเบส์ (Bayesian) 2) การประมาณค่าความสามารถของผู้สอบด้วยวิธีของเบส์ปรับใหม่ (Bayesian Updating) 3) การประมาณค่าความสามารถของผู้สอบด้วยวิธีความควรจะเป็นสูงสุด (Maximum Likelihood)

ขั้นตอนที่ 5 เกณฑ์การยุติการทดสอบ (Evaluate Termination Criterion) คือ การสิ้นสุดการทดสอบ เมื่อผู้สอบทำการทดสอบครบตามเงื่อนไข หรือเกณฑ์ที่กำหนด

4. ประโยชน์ของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์

Green (1983), Wainer (1990), Lalu and Chapman (2002) ได้กล่าวถึงประโยชน์ของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ไว้ สรุปได้ดังนี้

4.1 ช่วยเพิ่มประสิทธิภาพในการทดสอบโดยเฉพาะภาระงานการทดสอบต่างๆ ได้แก่ การดำเนินการทดสอบ การกำหนดเกณฑ์ และการให้คะแนน

4.2 แบบทดสอบที่เก็บไว้ในคลังข้อสอบในคอมพิวเตอร์จะมีความปลอดภัย (Security) ต่อการคัดลอกมากกว่า

4.3 ไม่มีข้อจำกัดเรื่องเวลาผู้สอบสามารถทำแบบทดสอบในเวลาที่แตกต่างกันตามความพร้อมของแต่ละบุคคลภายในช่วงเวลาที่กำหนด

4.4 ผู้สอบไม่เกิดความท้อแท้ในการทำข้อสอบเนื่องจากเป็นการคัดเลือกของข้อสอบที่เหมาะสมกับความสามารถของแต่ละบุคคล

4.5 การตรวจข้อสอบมีความชัดเจนขึ้น เนื่องจากไม่มีปัญหาที่เกิดจากกระดาษคำตอบ เช่น การทำเครื่องหมายไม่ชัดเจน ความไม่ชัดเจนในข้อที่เลือก เนื่องจากมีการลบเมื่อต้องการเปลี่ยนคำตอบเป็นต้น

4.6 ข้อสอบสามารถตรวจให้คะแนนและให้ผลย้อนกลับ (Feed Back) กับผู้สอบได้ทันที

4.7 สามารถคัดเลือกข้อสอบจากการตอบข้อสอบเบื้องต้นได้อย่างรวดเร็ว

4.8 ผู้สอบได้ทำข้อสอบที่มีความยากเหมาะสมกับระดับความสามารถ

4.9 มีการเสนอข้อสอบให้ผู้สอบตอบในรูปแบบที่หลากหลายทั้งภาพนิ่งภาพเคลื่อนไหวและเสียงประกอบ

5. ข้อจำกัดของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์

Latu and Chapman (2002) กล่าวถึงข้อจำกัดของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ไว้ดังนี้

5.1 ความปลอดภัยของแบบทดสอบ (Test Security) แม้ว่าจะมีนักวิจัยหลาย ๆ คน แสดงให้เห็นว่าการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์มีความปลอดภัยของแบบทดสอบ

มากกว่าการทำแบบทดสอบบนโต๊ะ (Desk Drawer) แต่ก็มีนักวิจัยบางท่านโต้แย้งว่าปัจจุบัน มีวิธีการขโมยข้อสอบได้หลายวิธี เช่น เพจเจอร์ (pager) กล้องขนาดเล็ก (Miniature Cameras) กล้องวิดีโอ (Video Transmitters) เครื่องบันทึกวิดีโอขนาดเล็ก (Micro Video Recorders) นอกจากนี้ผู้สอบยังสามารถคัดลอกข้อสอบจากเครื่องคอมพิวเตอร์ได้ง่าย และรวดเร็วกว่าการคัดลอกข้อสอบจากกระดาษคำตอบอีกด้วย

5.2 ปัญหาเกี่ยวกับผู้สอบ (Examinee Issues) แม้ว่าการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์จะมีประโยชน์ต่อผู้สอบมาก แต่ยังคงมีข้อจำกัดเกี่ยวกับผู้สอบ เช่น ผู้สอบหลายคนไม่เคยชินกับการใช้คอมพิวเตอร์ทำให้เกิดความวิตกกังวล ผู้สอบบางคนแสดงให้เห็นว่าการนำเสนอโดยใช้คอมพิวเตอร์เป็นสิ่งที่ยุ่งยาก (Difficult or Fatiguing) โดยเฉพาะการตอบโดยใช้เมาส์หรือคีย์บอร์ด

5.3 ค่าใช้จ่ายในการดำเนินการ (Financial Realities) การทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ มีค่าใช้จ่ายในการดำเนินการสร้างแบบทดสอบ การเตรียมการและการดำเนินการทดสอบที่ค่อนข้างสูง เช่น Educational Testing Service (ETS) ได้ศึกษาค่าใช้จ่ายในการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ พบว่ามีค่าใช้จ่ายที่ต้องจ่ายจริงมากกว่าวิธีการทดสอบแบบประเพณีนิยม ดังจะเห็นได้จากในสหรัฐอเมริกาผู้สอบจะต้องเสียค่าใช้จ่ายมากกว่า 100 ดอลลาร์ เพื่อทำการสอบ TOFFL ในระบบการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ ค่าใช้จ่ายที่แพงมากดังกล่าว มาจากการพัฒนาคลังข้อสอบการบริหารการทดสอบ และการพัฒนาโปรแกรมสำเร็จรูปซึ่งมีค่าใช้จ่ายในการพัฒนาและการนำไปใช้ในระบบที่สมบูรณ์ระหว่าง 3,000-5,000 ดอลลาร์

ตอนที่ 4 งานวิจัยเกี่ยวกับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์สำหรับการทดสอบวินิจฉัย

จากการศึกษาเอกสารงานวิจัยที่เกี่ยวข้องกับการทดสอบวินิจฉัยโดยใช้วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์

Tatsuoka and Tatsuoka (1997) ได้ใช้รูลสเปซโมเดล (Rule Space Model) ในการทดสอบแบบปรับเหมาะเพื่อพัฒนากระบวนการวินิจฉัย และการสอนซ่อมเสริม เรื่องปัญหาการบวกลบเศษส่วนโดยใช้คอมพิวเตอร์ กลุ่มตัวอย่างที่ใช้ได้แก่นักเรียนเกรด 7 และเกรด 8 ของโรงเรียนมัธยมศึกษาในเมืองเล็ก ๆ ในรัฐ Illinois ประเทศสหรัฐอเมริกาในปี ค.ศ. 1988 จำนวน 287 คน และในปี ค.ศ. 1989 จำนวน 191 คน การดำเนินการใช้กระบวนการที่ต่อเนื่องใน 2 ปี โดยเริ่มจากการทดสอบก่อนการซ่อมเสริม (Pretest) เพื่อจัดนักเรียนตามสถานะความรู้จำนวน 33 สถานะโดยพิจารณาจากการตอบคำถามและพิจารณาว่าใกล้กับสถานะความรู้ใดมากที่สุด การทดสอบจะหยุดสอบเมื่อผู้สอบถูกจัดอยู่ในสถานะความรู้ใดความรู้หนึ่งอย่างคงที่ (Stabilized on Some Point in Rule Space) หลังจากนั้น จึงจัดนักเรียนเพื่อรับการซ่อมเสริม โดยใช้เครื่องมือซ่อมเสริมในระบบ PLATO (A Computer-based Educational Teacher System Developed at University of Illinois) เลือกคุณลักษณะเพื่อใช้ในการซ่อมเสริมก่อนหลังโดยพิจารณาจากค่า Mahalanobis Distances ตามเส้นทางรูปต้นไม้ ซึ่งแสดงสถานะความรอบรู้ที่กำหนดไว้ เมื่อนักเรียนได้รับการฝึกหัดจนผ่านทุกคุณลักษณะในสถานะความรู้ที่ได้รับการซ่อมเสริมแล้ว นักเรียนจะได้รับการทดสอบภายหลังและการทดสอบซ้ำ (Posttest, Retention Test) ซึ่งดำเนินการทดสอบภายหลังจาก

การทดสอบภายหลัง 3 เดือน สำหรับนักเรียนที่ถูกจัดอยู่ในสถานะความรู้ที่ไม่ต้องได้รับการซ่อมเสริม ไม่ต้องทดสอบภายหลังและทดสอบซ้ำ ผลการวิจัยแสดงให้เห็นว่า รูลสเปซโมเดล (Rule Space Model) สามารถใช้ในการวินิจฉัยและการจัดเครื่องมือเพื่อการซ่อมเสริมได้อย่างมีประสิทธิภาพ

Grader (2004) ได้พัฒนาแบบสำรวจ Beck Depression Inventory (BDI) โดยใช้การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์เพื่อการวินิจฉัยและคัดแยกผู้ป่วยที่มีอาการซึมเศร้า โดยใช้เกณฑ์ช่วงความเชื่อมั่น (Confidence Interval: CI) ระดับความสามารถที่สูงกว่าหรือต่ำกว่าเกณฑ์คะแนนหรือเกณฑ์คะแนนจุดตัด เพื่อทดสอบประสิทธิภาพและประโยชน์ในการจำแนก Major Depression Episodes (MDE) และการวัดระดับความรุนแรงของความซึมเศร้า กลุ่มตัวอย่างที่ทำการทดสอบแบบสำรวจ Beck Depression Inventory (BDI) และแบบสำรวจ Structured Clinical Interview for DSM-IV Axis Disorder (SCID) จำนวน 744 คน และกลุ่มตัวอย่างที่เป็นผู้ป่วยทำแบบสำรวจ Hamilton Depression Rating Scale จำนวน 285 คน วิเคราะห์ข้อสอบโดยใช้ทฤษฎีการตอบสนองข้อสอบแบบตรวจให้คะแนนมากกว่า 2 ค่า โดยใช้ Graded Response Model (GRM) โดยการจำลองข้อมูล (Simulation Data) ซึ่งข้อสอบข้อแรกเป็นข้อที่ให้สารสนเทศสูงสุดของคะแนนความซึมเศร้าโดยเฉลี่ยของกลุ่มประชากร การเลือกข้อสอบข้อถัดไปพิจารณาจากข้อสอบที่ให้ค่าสารสนเทศสูงสุด ผลการวิจัยปรากฏว่า แบบสำรวจ BDI ที่ทดสอบโดยใช้วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์มีพื้นที่ใต้โค้ง (Area Under the Curve: AUC) ซึ่งเป็นตัวบ่งชี้ของแบบสำรวจ SCID ในการวินิจฉัย MDE มีค่าเท่ากับ 88% เท่า ๆ กับ BDI ฉบับเต็มแต่ใช้คำถามน้อยกว่า 5.6 กับ 21 ข้อ นอกจากนี้ยังชี้ให้เห็นว่า คะแนนความซึมเศร้าของ BDI ฉบับที่ทดสอบโดยใช้วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์มีค่าสัมประสิทธิ์สหสัมพันธ์กับคะแนนรวมของ BDI และคะแนนความซึมเศร้าของ Hamilton มีค่าเท่ากับ .92 และ .74 ตามลำดับ แสดงให้เห็นว่า แบบสำรวจ BDI ที่ทดสอบโดยใช้วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ มีประสิทธิภาพมากขึ้นโดยไม่ทำให้ความถูกต้องในการประเมินผลลดน้อยลง

Guzman and Conejo (2004) ได้พัฒนาโมเดลสำหรับใช้ในการเลือกคำตอบแบบมากกว่า 2 ค่า (Polytomous) สำหรับใช้ในการวินิจฉัยผู้เรียนใน Intelligent Tutoring Systems (ITS) โดยใช้วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ซึ่งเป็นแบบทดสอบความสมดุลเนื้อหา (Content Balanced Test) ประมาณค่าความรู้ของผู้เรียนโดยใช้การโค้งกระจายความน่าจะเป็นของข้อสอบที่ผู้เรียนทำ ประมาณค่าโดยวิธีของเบส์ (Bayes theorem) คัดเลือกข้อสอบข้อถัดไปโดยใช้ค่าความแปรปรวนของการกระจายความรู้ที่คาดหวังของผู้เรียนที่มีค่าน้อยที่สุด (Minimizes the Expectation of Posteriori Student Knowledge Distribution Variance) วิเคราะห์ค่าพารามิเตอร์ของข้อสอบโดยใช้ทฤษฎีการตอบสนองข้อสอบแบบตรวจให้คะแนนมากกว่า 2 ค่า ด้วย Partial Credit Model โดยการจำลองข้อมูล (Simulation Data) เพื่อเปรียบเทียบจำนวนเฉลี่ยของข้อสอบที่ใช้ค่าเฉลี่ยของความคลาดเคลื่อนในการประมาณค่าความรู้ของผู้สอบและร้อยละของความถูกต้องในการประเมินผู้เรียน สำหรับข้อสอบที่มีค่าอำนาจจำแนกแตกต่างกันระหว่างการวินิจฉัยที่ใช้การเลือกตอบแบบ 2 ค่า (Dichotomous) และการเลือกตอบแบบมากกว่า 2 ค่า (Polytomous) ผลจากการวิจัยแสดงให้เห็นว่า การเลือกข้อสอบแบบมากกว่า 2 ค่า (Polytomous) ให้ผลดีกว่าการเลือกคำตอบแบบ 2 ค่า (Dichotomous) สำหรับข้อสอบที่มีการกระจายต่ำ ทั้ง

จำนวนข้อสอบที่ใช้ ค่าความคลาดเคลื่อนในการประมาณค่า ความรู้ของนักเรียนและความถูกต้องในการประเมิน โมเดลที่พัฒนาขึ้นให้ข้อมูลการตอบของนักเรียนได้มากกว่าโมเดลการเลือกตอบแบบ 2 ค่า และให้ผลการประเมินที่ถูกต้องมากกว่าโดยใช้ข้อคำถามน้อยลง

McGlohen and Chang (2008) ได้ประยุกต์วิธีการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ในการประเมินเชิงวินิจฉัยทางปัญญา (Cognitive Diagnosis Assessment) โดยมุ่งเน้นการให้ข้อมูลย้อนกลับเชิงวินิจฉัยจากเวกเตอร์คุณลักษณะเป็นรายบุคคลในแต่ละบุคคล (Individual Attribute Vector) ที่เชื่อมโยงกับเทคโนโลยีการทดสอบแบบใหม่ ซึ่งช่วยให้ครูและผู้เรียนได้รับประโยชน์จากกระบวนการทดสอบ โดยใช้วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์เป็นแบบ Shadow Testing คัดเลือกข้อสอบโดยวิธี Minimization of Shannon Entropy ซึ่งเป็นยุทธวิธีในการทดสอบความไม่แน่นอนที่มีค่าต่ำสุด (Minimal Uncertainly) และยุทธวิธี Maximization of Kullback Leibler Information (K-L) ซึ่งเป็นยุทธวิธีที่ใช้ในการคัดเลือกข้อสอบและเป็นดัชนีในกระบวนการคัดเลือกข้อสอบในการประเมินเชิงวินิจฉัย โดยใช้ค่าสารสนเทศ KL ที่มีค่าสูงสุด ออกแบบการเลือกข้อสอบเป็น 3 กรณีคือ กรณีที่ 1 เลือกข้อสอบจากการประมาณค่าความสามารถ (Theta) เท่านั้น กรณีที่ 2 เลือกข้อสอบจากการประมาณค่าเวกเตอร์คุณลักษณะ ($\propto$) เท่านั้น และกรณีที่ 3 เลือกข้อสอบทั้งจากการประมาณค่าระดับความสามารถและการประมาณค่าเวกเตอร์คุณลักษณะ โดยคัดเลือกกลุ่มตัวอย่างด้วยวิธีการสุ่มอย่างง่าย จากการประเมินระดับกว้างของรัฐ (State-mandated Large-scale Assessment) ได้กลุ่มตัวอย่างจำนวน 2,000 คน คลังข้อสอบประกอบด้วยข้อสอบวิชาคณิตศาสตร์ 396 ข้อ วิชาการอ่าน 324 ข้อ การประมาณค่าพารามิเตอร์ของผู้สอบทั้งคะแนนและแบบแผนความรู้ของคุณลักษณะ (Attribute Mastery Patterns) โดยใช้ทฤษฎีการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ และโมเดล Fusion Model ด้วยโปรแกรม Bilog-MG และ Arpeggio 1.2 ข้อมูลที่ใช้เป็นข้อมูลจากการจำลอง (Simulation Data) แบบการแทนค่า (Replacement) จากคะแนนจริง จากกลุ่มตัวอย่าง 6,000 คน และ 3,000 คน ตามลำดับ ผลการวิจัยชี้ให้เห็นว่า การใช้การประมาณค่าระดับความสามารถให้ความถูกต้องในการประมาณค่าได้มากที่สุด และข้อสอบมีอัตราการแสดงข้อสอบซ้ำต่ำ (Minimal Item Exposure Rates) ให้การประมาณค่าได้อย่างถูกต้อง

Rapp and Templin (2008) ได้ศึกษาเกี่ยวกับอิทธิพลของการกำหนดคุณสมบัติของ Q-matrix ผิดพลาดที่มีต่อการประมาณค่าพารามิเตอร์และอัตราการกำหนดคุณสมบัติผิดพลาดสำหรับ DINA Model ซึ่งอยู่ในโมเดลลำดับขั้นคุณลักษณะแฝงสำหรับการจำแนกประเภทของคำตอบแบบหลายคำตอบ เป็นการศึกษาจำลองโดยใช้ Q-matrix เป็นแนวทางการประเมินคุณลักษณะทั้งหมด 15 แบบ จากตัวแปรอิสระ 4 ตัว โดยการเปลี่ยนคุณสมบัติที่ผิดพลาดของข้อสอบแต่ละข้อจาก “0” เป็น “1” หรือจาก “1” เป็น “0” ผลการวิจัยชี้ให้เห็นว่า ส่วนประกอบที่เป็นคุณลักษณะเฉพาะของข้อสอบ มีอิทธิพลทำให้การประมาณค่าเกินความเป็นจริงและค่าพารามิเตอร์มีความผิดพลาดเมื่อคุณลักษณะถูกลบออกจาก Q-matrix และมีการประมาณค่าพารามิเตอร์การเดาสูงเกินความเป็นจริงเมื่อมีการเพิ่มคุณลักษณะให้กับ Q-matrix การจำแนกผิดมีอัตราสูงสำหรับลำดับขั้นคุณลักษณะที่มีกลุ่มคุณลักษณะถูกลบออกจาก Q-matrix

จากงานวิจัยดังกล่าวข้างต้น จะเห็นได้ว่าการพัฒนาวิธีการวินิจฉัยโดยใช้แนวคิดโมเดลลำดับขั้นคุณลักษณะในการกำหนดคุณลักษณะหลักของเนื้อหาให้มีลักษณะเฉพาะก่อนการพัฒนาชุดข้อสอบ และการวินิจฉัยโดยใช้วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ยังดำเนินการไม่แพร่หลาย โดยเฉพาะในประเทศไทยยังไม่พบเอกสารงานวิจัยเกี่ยวกับวิธีการวินิจฉัยทักษะการอ่านภาษาอังกฤษ โดยใช้โมเดลลำดับขั้นคุณลักษณะ และการวินิจฉัยโดยใช้วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ดังนั้นการประยุกต์แนวคิดโมเดลลำดับขั้นคุณลักษณะและการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์สำหรับวิธีการวินิจฉัยคงต้องมีการดำเนินการที่หลากหลาย เพื่อให้ได้ข้อสรุปของผลการวินิจฉัยที่ชัดเจนยิ่งขึ้น

ผู้วิจัยจึงพัฒนาวิธีการวินิจฉัยทักษะการอ่านภาษาอังกฤษ โดยใช้โมเดลลำดับขั้นคุณลักษณะ และการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ เพื่อให้ได้แบบทดสอบวินิจฉัยที่มีการจำแนกเนื้อหาออกเป็นลำดับขั้น ก่อนการสร้างชุดข้อสอบ และสามารถนำไปใช้ในการทดสอบวินิจฉัยทักษะการอ่านภาษาอังกฤษของนักศึกษาระดับปริญญาตรี โดยใช้วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ เพื่อให้ได้สารสนเทศที่เป็นความรอบรู้หรือข้อบกพร่องในตัวผู้เรียนเป็นรายบุคคลได้ชัดเจนขึ้น ซึ่งจะเป็นประโยชน์สำหรับอาจารย์ผู้สอนภาษาอังกฤษสามารถนำผลไปใช้ในการวินิจฉัยการเรียนรู้ของนักศึกษาได้รวดเร็ว ช่วยลดเวลาในการวินิจฉัยนักศึกษาเป็นรายบุคคล รวมทั้งสามารถสรุปผลและให้ข้อมูลย้อนกลับแก่นักศึกษาในส่วนที่เป็นความรอบรู้หรือจุดบกพร่องในทันที สำหรับนักศึกษาได้นำไปพัฒนาตนเอง เพื่อเพิ่มศักยภาพการเรียนรู้ในทักษะการอ่านภาษาอังกฤษของนักศึกษาในระดับสูงต่อไป

มหาวิทยาลัยบูรพา Burapha University