

การทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์

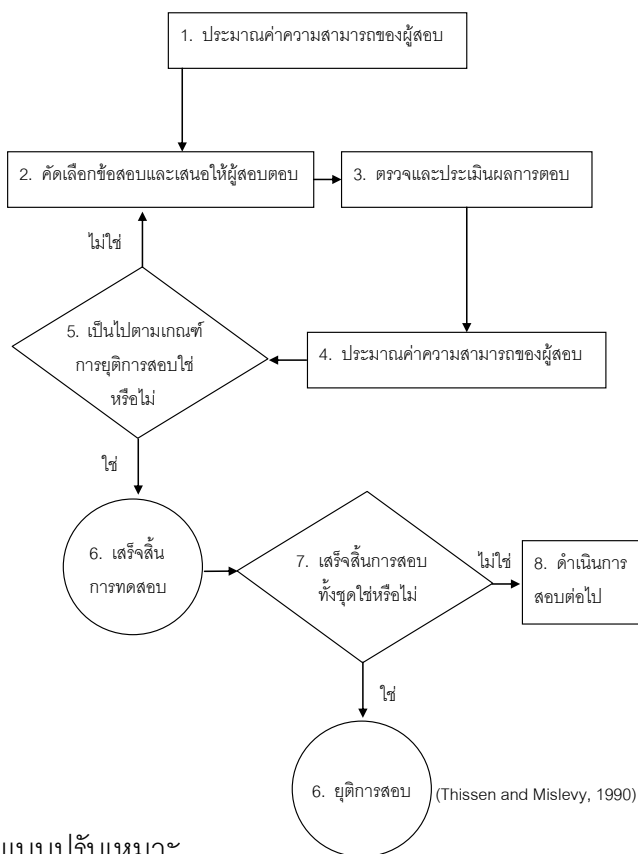
ดร.อมรรัตน์ สร้อยสังวาลย์*

การทดสอบแบบปรับเหมาะ (adaptive testing or tailored testing) เป็นการทดสอบที่มุ่งวัดคุณลักษณะที่ต้องการ โดยการคัดเลือกข้อสอบตามความสามารถของผู้สอบในระหว่างกระบวนการทดสอบ การคัดเลือกข้อสอบจะพิจารณาจากผลการตอบข้อสอบหรือข้อคำถามที่มีมาก่อน ถ้าตอบข้อสอบข้อนั้นได้ถูกต้อง ข้อสอบข้อต่อไปก็จะยากขึ้น แต่ถ้าตอบข้อสอบข้อนั้นผิดข้อสอบข้อต่อไปก็จะง่ายลง (Weiss, 1983; Weiss and Schleisman, 1999) แบบทดสอบที่ใช้ในการทดสอบแบบปรับเหมาะจึงเป็นแบบทดสอบที่ประกอบไปด้วยชุดของข้อสอบที่แตกต่างกัน สำหรับการสอบของแต่ละบุคคลซึ่งเป็นลักษณะที่ตรงกันข้ามกับแบบทดสอบแบบประเพณีนิยม (conventional tests) ที่ผู้สอบทุกคนจะถูกสอบด้วยแบบทดสอบชุดเดียวกัน Weiss (1988, 1990)

แบบทดสอบแบบปรับเหมาะฉบับแรก ได้แก่ แบบทดสอบเชาว์ปัญญาของบินเนท์ (Binet intelligence test) พัฒนาขึ้นโดย Alfred Binet นักจิตวิทยาชาวฝรั่งเศส แบบทดสอบเชาว์ปัญญาของบินเนท์ที่ใช้การเลือกชุดของข้อสอบเบื้องต้นตามระดับอายุ (ความยาก) ข้อสอบจะถูกให้คะแนนและถูกเลือกให้ผู้สอบทำตามทางเลือกที่กำหนดไว้คงที่ (fixed branching) โดยผู้บริหารการสอบ เมื่อผู้สอบทำข้อสอบในระดับอายุจริงได้ไม่ถูกต้องก็จะถูกให้ทำข้อสอบในระดับอายุที่ต่ำกว่า แต่ถ้าทำได้ถูกต้องก็จะถูกให้ทำข้อสอบในระดับอายุที่สูงกว่า จนกระทั่งถึงระดับเพดาน (Ceiling level) เมื่อผู้สอบทำข้อสอบทุกข้อไม่ถูกต้องหรือระดับพื้นฐาน (basal level) เมื่อผู้สอบทำข้อสอบทุกข้อได้ถูกต้อง เกณฑ์ที่ใช้ในการยุติการสอบจะแตกต่างกันตามระดับความสามารถของแต่ละบุคคลเมื่อสามารถระบุระดับเพดานและระดับพื้นฐานของผู้สอบได้ (Weiss, 1983; Weiss, 1988, 1990) แม้ว่าแบบทดสอบเชาว์ปัญญาของบินเนท์จะถูกนำไปใช้อย่างแพร่หลายและเป็นต้นฉบับในการกำหนดหลักการของการทดสอบแบบปรับเหมาะในปัจจุบัน แต่แบบทดสอบเชาว์ปัญญาของบินเนท์ก็มีข้อจำกัดที่ผู้สอบอาจแสดงผลการตอบไม่คงเส้นคงวาระหว่างการทดสอบ ผู้สอบบางคนอาจทำข้อสอบบางข้อได้ดีแตกต่างกันเนื่องจากอิทธิพลของเชื้อชาติ เพศ วัฒนธรรม รวมทั้งยังมีค่าใช้จ่ายในการทดสอบสูงและใช้เวลาในการสอบมาก (ศิริชัย กาญจนวาสี, 2538, 2545) แบบทดสอบแบบปรับเหมาะอื่น ๆ เริ่มพัฒนาขึ้นปี ค.ศ. 1950 โดยเริ่มตั้งแต่การเขียนตอบลงบนกระดาษคำตอบ (paper-and-pencil) จนถึงการใช้คอมพิวเตอร์ช่วยในการทดสอบ (computer assisted testing) ซึ่งมีการพัฒนาวีธีเลือกข้อสอบเริ่มต้นและกำหนดเกณฑ์ในการยุติการสอบที่แตกต่างกัน แต่ยังคงใช้หลักการที่เน้นความเหมาะสมกับความสามารถของผู้สอบเช่นเดียวกับแบบทดสอบเชาว์ปัญญาของบินเนท์ และมักจะเกี่ยวข้องกับคำว่า เหมาะสม (tailored) ความไม่แน่นอนในการตอบ (response-contingent) โปรแกรม (programmed) คอมพิวเตอร์ (computerized) อัตโนมัติ (automated) แต่ละบุคคล (individualized) ทางแยก (branched) และลำดับขั้นของการทดสอบ (sequential testing) (Weiss, 1988, 1990)

*อาจารย์ประจำคณะครุศาสตร์ มหาวิทยาลัยราชภัฏเพชรบุรี

การดำเนินการทดสอบแบบปรับเหมาะจะเกี่ยวข้องกับคำถามสำคัญ 3 คำถาม ได้แก่ จะเลือกข้อสอบที่ใช้ในการเริ่มต้นการสอบได้อย่างไร (How to START) จะเลือกข้อสอบข้อถัดไปได้ได้อย่างไร (How to CONTINUE) และรู้ได้อย่างไรว่าจะยุติการสอบ (How to STOP) (Thissen and Mislevy, 1990; Wainer, 1990) คำถามดังกล่าวจะไม่พบในการทดสอบแบบประเพณีนิยมเนื่องจากในการเริ่มต้นทำแบบทดสอบมักจะเริ่มจากการทำข้อสอบข้อที่ 1 ข้อต่อ ๆ ไป จนถึงข้อสุดท้าย แต่ในการทดสอบแบบปรับเหมาะการเลือกข้อสอบที่ใช้ในการประมาณค่าความสามารถเบื้องต้นหรือเป็นขั้นตอนของการเริ่มการทดสอบเป็นสิ่งสำคัญและมีผลต่อการประมาณค่าความสามารถของผู้สอบเพื่อการเลือกข้อสอบข้อต่อไป รวมถึงการยุติการสอบและการประมาณค่าความสามารถสุดท้ายของผู้สอบ การเริ่มต้นการสอบ การดำเนินการสอบและการยุติการสอบจึงเป็นสิ่งที่ต้องพิจารณาเพื่อให้การทดสอบแบบปรับเหมาะมีประสิทธิภาพสูงสุดและมักจะเป็นประเด็นที่ใช้ในการศึกษาวิจัยพัฒนาเกี่ยวกับการทดสอบแบบปรับเหมาะ โดยทั่วไปการทดสอบแบบปรับเหมาะมักดำเนินการโดยเริ่มจากการประมาณค่าความสามารถของผู้สอบ คัดเลือกข้อสอบที่เหมาะสมกับความสามารถของผู้สอบและเสนอให้ผู้สอบทำ หลังจากนั้นจึงประมาณค่าความสามารถของผู้สอบอีกครั้งหนึ่ง ตรวจและประเมินคำตอบ พิจารณาว่าเป็นไปตามเกณฑ์ที่ใช้ในการยุติการสอบใช่หรือไม่ ถ้าไม่ใช่ก็ดำเนินการคัดเลือกข้อสอบข้อต่อไป ถ้าใช่ก็สิ้นสุดการสอบสำหรับแบบทดสอบฉบับนั้นและพิจารณาว่าเสร็จสิ้นการทดสอบทุกฉบับในแบบทดสอบชุดนั้นแล้วใช่หรือไม่ ถ้าไม่ใช่ก็ดำเนินการสอบแบบทดสอบฉบับต่อไป ถ้าใช่ก็ยุติการสอบ (Thissen and Mislevy, 1990) ดังแสดงในแผนภูมิที่ 1



แผนภูมิที่ 1 ขั้นตอนการทดสอบแบบปรับเหมาะ

ประเภทของการทดสอบแบบปรับเหมาะ

การจำแนกประเภทของการทดสอบแบบปรับเหมาะพิจารณาได้จากเครื่องมือที่ใช้ในการทดสอบ และยุทธวิธีที่ใช้ในการทดสอบ (ศิริชัย กาญจนวาสี, 2538, 2545) มีรายละเอียด ดังนี้

1. การทดสอบแบบปรับเหมาะจำแนกตามเครื่องมือที่ใช้

การทดสอบแบบปรับเหมาะจำแนกตามเครื่องมือที่ใช้ได้ 2 ประเภท ประเภทแรกเป็นการทดสอบแบบปรับเหมาะโดยไม่ใช้คอมพิวเตอร์และประเภทที่ 2 เป็นการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ การทดสอบแบบปรับเหมาะโดยไม่ใช้คอมพิวเตอร์ ประกอบด้วย แบบทดสอบเซวาร์นบัญญัติของบิเนท์ แบบทดสอบเลือกคำตอบเป็นกลุ่มแบบสองขั้นตอน (Two-Stage Test) แบบทดสอบปรับระดับความสามารถที่ยืดหยุ่น (Flexilevel Test) และแบบทดสอบกำหนดตามการแยกทาง (Branching Test) ส่วนการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ซึ่งเป็นการนำคอมพิวเตอร์มาใช้เพื่อความสะดวกในการคัดเลือกข้อสอบและประมาณค่าความสามารถของผู้สอบ เริ่มพัฒนาขึ้นประมาณปี ค.ศ. 1970 (Weiss, 1983)

2. การทดสอบแบบปรับเหมาะจำแนกตามยุทธวิธีที่ใช้ในการทดสอบ

การทดสอบแบบปรับเหมาะจำแนกตามยุทธวิธีที่ใช้ในการทดสอบได้ 2 ประเภท ประเภทแรกเป็นการทดสอบแบบปรับเหมาะโดยใช้ยุทธวิธีสองขั้นตอน (Two-stage Strategies) และประเภทที่ 2 เป็นการทดสอบแบบปรับเหมาะโดยใช้ยุทธวิธีหลายขั้นตอน (Multi-stage Strategies) ประกอบด้วยยุทธวิธีหลายขั้นตอนแบบทางแยกคงที่ (Fixed-Branching) และยุทธวิธีหลายขั้นตอนแบบทางแยกแปรผัน (Variable Branching)

ยุทธวิธีที่ใช้ในการทดสอบแบบปรับเหมาะ

ยุทธวิธีที่ใช้ในการทดสอบแบบปรับเหมาะประกอบด้วย ยุทธวิธีสองขั้นตอน (Two-stage Strategies) และยุทธวิธีหลายขั้นตอน (Multi-stage Strategies) มีรายละเอียด ดังนี้

1. ยุทธวิธีสองขั้นตอน (Two-stage Strategies)

ในปี ค.ศ. 1970 และ ปี ค.ศ. 1980 นักวิจัยได้มีความพยายามในการพัฒนาแบบทดสอบแบบปรับเหมาะโดยเน้นไปที่ขั้นตอนการคัดเลือกข้อสอบ เพื่อให้ได้จำนวนข้อสอบที่น้อยที่สุดในการประมาณค่าความสามารถของผู้ตอบได้อย่างถูกต้อง ยุทธวิธีสองขั้นตอนจึงได้รับการพัฒนาขึ้นเพื่อตอบสนองวัตถุประสงค์ดังกล่าว โดยใช้แบบทดสอบสองฉบับในการทดสอบสองขั้นตอน แบบทดสอบแต่ละฉบับในแต่ละขั้นตอนจะประกอบไปด้วยกลุ่มของข้อสอบที่มุ่งวัดเนื้อหาใดเนื้อหาหนึ่งซึ่งกำหนดทิศทางคำตอบสำหรับผู้สอบไว้ เรียกว่า testlets (Thissen and Mislevy, 1990) แบบทดสอบในแต่ละขั้นตอนมีรายละเอียด ดังนี้

1.1 แบบทดสอบขั้นต้นแรกหรือแบบทดสอบเพื่อกำหนดทิศทาง (first-stage test or routing test)

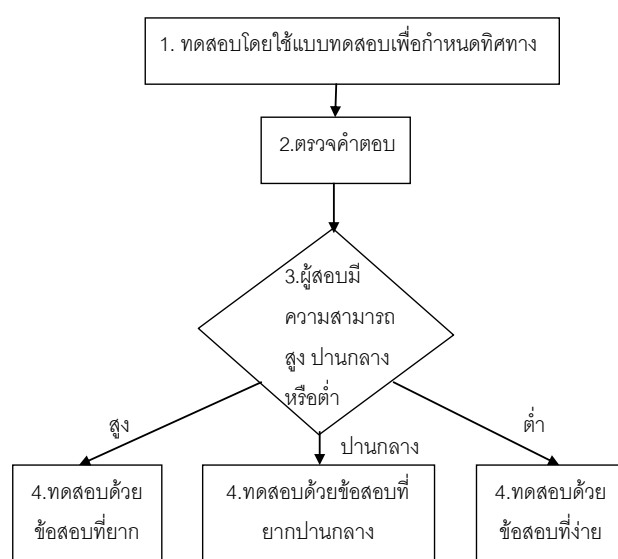
แบบทดสอบขั้นต้นแรกหรือแบบทดสอบเพื่อกำหนดทิศทางเป็นแบบทดสอบที่ค่อนข้างสั้น ประกอบด้วยข้อสอบที่มีระดับความยากกระจายในช่วงกว้างหรือมีระดับความยากง่ายปานกลาง เพื่อให้ผู้สอบทำในขั้นต้นแรกและตรวจข้อสอบทันทีเพื่อใช้ผลการทดสอบเป็นเกณฑ์ในการคัดเลือกแบบทดสอบในขั้นที่สองโดยพิจารณาตามค่าความยากง่ายของแบบทดสอบที่เหมาะสมกับระดับความสามารถของผู้สอบซึ่งประมาณได้จากการทดสอบแบบทดสอบฉบับแรก (นันทิยา พึ่งคำ, 2531; ศิริชัย กาญจนวาสี, 2538, 2545)

1.2 แบบทดสอบขั้นตอนที่สอง แบบทดสอบหลักหรือแบบทดสอบวัดผล (second-stage test, main test or measurement test)

แบบทดสอบขั้นตอนที่สอง แบบทดสอบหลักหรือแบบทดสอบวัดผลเป็นแบบทดสอบที่ถูกคัดเลือกให้เหมาะสมกับความสามารถของผู้สอบซึ่งประมาณได้จากการทดสอบในขั้นแรกมักเป็นแบบทดสอบที่ยาวกว่าแบบทดสอบในขั้นแรกและประกอบด้วยแบบทดสอบย่อย ประมาณ 3-5 ฉบับ แยกตามระดับความยาก ในแต่ละฉบับจะมีข้อสอบจำนวน 20-30 ข้อ (นันทิยา พึ่งคำ, 2531; สายชล อบรม, 2539) คะแนนที่ได้จากแบบทดสอบแต่ละฉบับไม่สามารถเปรียบเทียบกันได้โดยตรง เนื่องจากข้อสอบแต่ละฉบับมีความยากง่ายต่างกันวิธีนำคะแนนมาเปรียบเทียบกันอาจทำได้โดยการนำคะแนนของแต่ละฉบับมาถ่วงด้วยค่าความยากเฉลี่ยของแต่ละฉบับเพื่อปรับค่าความยากให้เท่ากัน (ศิริชัย กาญจนวาสี, 2538, 2545)

การคัดเลือกข้อสอบและจัดชุดของข้อสอบในการทดสอบทั้งสองขั้นต้นดังกล่าว ศิริชัย กาญจนวาสี (2538, 2545) ได้เสนอไว้ว่าสามารถนำทฤษฎีการตอบสนองข้อสอบ (Item Response Theory) มาประยุกต์ใช้ได้เพื่อให้การประมาณค่าความสามารถของผู้สอบไม่ขึ้นกับระดับความยากของแบบทดสอบ

ขั้นตอนการทดสอบโดยใช้ยุทธวิธีสองขั้นต้น แสดงได้ในแผนภูมิที่ 2



แผนภูมิที่ 2 ขั้นตอนการทดสอบแบบปรับเหมาะโดยใช้ยุทธวิธีสองขั้นต้น

สำหรับข้อดีและข้อจำกัดของการทดสอบโดยใช้ยุทธวิธีสองขั้นตอน ศิริชัย กาญจนวาสี (2538, 2545) ได้กล่าวไว้ว่า มีข้อดีที่ช่วยลดจำนวนข้อสอบที่แต่ละคนจะต้องทำ โดยไม่ทำให้ความแม่นยำในการวัดลดลง แต่มีข้อจำกัดกรณีทดสอบกับคนจำนวนมากในเวลาเดียวกัน การตรวจข้อสอบกำหนดทิศทางจะต้องรีบเร่งตรวจและใช้คนจำนวนมาก รวมทั้งการปรับเหมาะของกระบวนการทดสอบเกิดขึ้นเพียงครั้งเดียวหลังจากการทดสอบในขั้นแรก ซึ่งถ้าเกิดความคลาดเคลื่อนในขั้นนี้อาจทำให้เกิดความคลาดเคลื่อนในการการทดสอบในขั้นที่สองได้ ความคลาดเคลื่อนในการจำแนกกลุ่มความสามารถของผู้สอบอาจเกิดขึ้นได้ประมาณร้อยละ 20ตลอดจนแบบทดสอบที่มีช่วงค่าความยากที่แคบจะทำให้ได้สารสนเทศค่อนข้างน้อย แต่ถ้าสร้างแบบทดสอบวัดผลในขั้นที่สองให้มีช่วงความยากที่ค่อนข้างกว้างก็จะขัดแย้งกับหลักการของการทดสอบแบบปรับเหมาะ ทั้งนี้ได้เสนอแนะการแก้ปัญหาดังกล่าวไว้ว่าสามารถดำเนินการได้โดยใช้คอมพิวเตอร์ในการทดสอบแบบปรับเหมาะ

2. ยุทธวิธีหลายขั้นตอน (Multi-stage Strategies)

การทดสอบแบบปรับเหมาะโดยใช้ยุทธวิธีหลายขั้นตอนเป็นการใช้การทดสอบที่มากกว่า 2 ขั้นตอน ในการกำหนดทางแยก (branch) ระหว่างการทดสอบ จำแนกได้เป็น รูปแบบการทดสอบแบบทางแยกคงที่และรูปแบบการทดสอบแบบทางแยกแปรผัน (Hambleton and Swaminathan, 1985) มีรายละเอียดดังนี้

2.1 รูปแบบทางแยกคงที่ (Fixed-branching model)

รูปแบบทางแยกคงที่ที่ใช้ในยุทธวิธีหลายขั้นตอนเหมือนยุทธวิธีสองขั้นตอนในส่วนของการกำหนดทางแยกที่คงที่ แต่ต่างกันในส่วนของการตัดสินใจแยกทางซึ่งในการใช้ยุทธวิธีสองขั้นตอนกำหนดการตัดสินใจเพียงครั้งเดียวจากการทำแบบทดสอบกำหนดทิศทางในขั้นแรกไปการทำแบบทดสอบวัดผลในขั้นที่สอง ส่วนรูปแบบทางแยกคงที่ในการใช้ยุทธวิธีหลายขั้นตอนจะกำหนดการตัดสินใจหลังจากที่ผู้สอบทำข้อสอบในแต่ละข้อตามโครงสร้างของแบบทดสอบที่กำหนดไว้ล่วงหน้า โดยคำนวณค่าพารามิเตอร์ของข้อสอบไว้แล้ว ผู้สอบแต่ละคนจะได้ทำแบบทดสอบที่มีโครงสร้างของแบบทดสอบเดียวกันแต่จะได้ทำข้อสอบในข้อที่แตกต่างกันตามกฎการคัดเลือกข้อสอบที่กำหนดไว้รูปแบบในการกำหนดโครงสร้างของแบบทดสอบ สรุปได้ดังนี้

1) รูปแบบปิรามิดหรือรูปแบบโครงสร้างต้นไม้ (Pyramidal model) เป็นรูปแบบที่มีการวิจัยและการนำไปใช้จำนวนมาก มีการจัดโครงสร้างของข้อสอบเหมือนรูปทรงของปิรามิดหรือต้นไม้ โดยมีรูปแบบการกำหนดค่าความยาก วิธีการแยกทางและจำนวนข้อสอบที่แตกต่างกัน ประกอบด้วย แบบขนาดความยากคงที่ (constant step size) แบบขนาดความยากแปรผัน (variable step size) แบบปิรามิดข้างตัด (truncated pyramid) แบบปิรามิดที่มีหลายข้อในแต่ละขั้น (multiple pyramids) และแบบปิรามิดที่ให้ค่าน้ำหนักตัวเลือกเพื่อแยกทาง (differential response option branching) การทดสอบจะดำเนินการตามโครงสร้างของแบบทดสอบและข้อกำหนดในการแยกทางที่กำหนดไว้

2) รูปแบบปรับระดับแบบยืดหยุ่นหรือรูปแบบเฟล็กซ์เลเวล (Flexilevel model) เป็นรูปแบบที่ Frederic Lord พัฒนาขึ้นเมื่อปี ค.ศ. 1971 ในแบบทดสอบชื่อ self-scoring flexilevel test (Thissen and Misselevy, 1990) ซึ่งเป็นแบบทดสอบที่ประกอบด้วยข้อสอบที่มีช่วงห่างของค่าความยากเท่ากัน โดยจัดเรียงข้อสอบจากข้อที่ง่ายที่สุดไปข้อที่ยากที่สุด การทดสอบจะเริ่มจากการให้ผู้สอบทำข้อสอบที่มีค่าความยากปานกลาง ถ้าผู้สอบตอบข้อสอบได้ถูกต้องข้อสอบข้อต่อไปจะยากขึ้น แต่ถ้าผู้สอบตอบข้อสอบได้ไม่ถูกต้องข้อต่อไปจะเป็นข้อที่ง่ายลง การทดสอบจะดำเนินการไปจนกระทั่งผู้สอบทำแบบทดสอบได้จำนวนครึ่งหนึ่งของแบบทดสอบก็จะยุติการทดสอบ ในระยะแรกที่มีการพัฒนา ดำเนินการทดสอบโดยใช้การตอบบนกระดาษคำตอบ (paper-and-pencil) การเตรียมแบบทดสอบดำเนินการโดยการเรียงข้อสอบจากข้อที่ง่ายที่สุดไปข้อที่ยากที่สุดและแบ่งข้อสอบออกเป็นสองกลุ่ม ได้แก่ กลุ่มข้อสอบง่ายและกลุ่มข้อสอบยาก โดยมีข้อสอบที่มีค่าความยากง่ายปานกลางจำนวนหนึ่งข้อ ข้อสอบในแต่ละกลุ่มจะถูกจัดพิมพ์แยกในแต่ละส่วนของหน้ากระดาษ (กลุ่มข้อสอบง่ายอยู่ด้านซ้าย กลุ่มข้อสอบยากอยู่ด้านขวา) ในการทดสอบจะใช้การทำสัญลักษณ์ (marking bubbles) สำหรับข้อที่ตอบถูกและข้อที่ตอบผิดแตกต่างกันโดยการใช้สี (scratch-off) ข้อที่ตอบถูกจะทำเครื่องหมายจุดด้วยสีน้ำเงิน ส่วนข้อที่ตอบผิดจะทำเครื่องหมายจุดด้วยสีแดง การทดสอบเริ่มจากการให้ผู้สอบทำข้อสอบที่มีความยากง่ายปานกลางถ้าตอบผิดก็เลือกข้อสอบข้อต่อไปในกลุ่มของข้อสอบง่ายซึ่งอยู่ด้านซ้ายของกระดาษ ถ้าตอบข้อดังกล่าวถูกต้องก็ทำเครื่องหมายจุดสีน้ำเงินและเลือกข้อสอบข้อต่อไปในกลุ่มข้อสอบยากซึ่งอยู่ด้านขวาของกระดาษ แต่ถ้าตอบผิดก็ทำเครื่องหมายจุดสีแดงในข้อนั้นและเลือกข้อสอบข้อต่อไปที่อยู่ในกลุ่มข้อสอบง่ายซึ่งอยู่ด้านซ้ายของกระดาษ ดำเนินการจนกระทั่งทำข้อสอบได้ครึ่งหนึ่งของแบบทดสอบจึงยุติการสอบ

3) รูปแบบปรับระดับขั้น (Stradaptive model) เป็นรูปแบบที่ David J. Weiss พัฒนาขึ้นในปี ค.ศ. 1973 ในแบบทดสอบชื่อ Stratified adaptive (stradaptive) test โดยมีแนวคิดว่าคุณคนมีความสามารถที่แปรผันได้อย่างเป็นอิสระในการทำแบบทดสอบฉบับหนึ่ง ๆ ความคงเส้นคงวาในการทำแบบทดสอบ (consistency) จึงเป็นสิ่งสำคัญที่ควรพิจารณา (Weiss, 1983) แบบทดสอบแบบปรับระดับขั้นพัฒนามาจากแบบทดสอบแบบปิรามิด แต่แตกต่างกันที่แบบทดสอบแบบปิรามิดมีข้อสอบที่มีค่าความยากระดับเดียว ส่วนแบบทดสอบแบบปรับระดับขั้นมีข้อสอบที่มีค่าความยากหลายระดับ (Hambleton and Swaminathan, 1985) ค่าความยากของข้อสอบจะถูกแบ่งเป็นระดับขั้น (stratified) แต่ละระดับขั้นประกอบด้วยข้อสอบที่มีค่าความยากอยู่ในช่วงที่กำหนดเดียวกัน ในการทดสอบมีการดำเนินการคล้ายกับการทดสอบโดยใช้แบบทดสอบเชวาร์นีย์ บัญญาของบิเนท์ โดยผู้สอบจะเริ่มทำข้อสอบข้อแรกที่แตกต่างกันตามระดับความสามารถเบื้องต้น หลังจากนั้นก็จะประมาณค่าความสามารถของผู้สอบจากข้อสอบข้อแรกเพื่อคัดเลือกข้อสอบข้อต่อไป ถ้าผู้สอบตอบข้อสอบข้อแรกถูกข้อสอบข้อต่อไปก็จะเป็นข้อที่ยากขึ้น แต่ถ้าตอบผิดข้อต่อไปจะเป็นข้อที่ง่ายลง การทดสอบจะดำเนินการไปจนกระทั่งผู้สอบทำข้อสอบในระดับนั้นได้ไม่ถูกต้องทุกข้อหรือจนถึงระดับเพดาน (ceiling stratum) หรือทำคะแนนได้ไม่เกินคะแนนที่ได้จากการเดา (chance score)

2.2 รูปแบบทางแยกแปรผัน (Variable-branching model)

รูปแบบทางแยกแปรผันเป็นรูปแบบการทดสอบที่ไม่ได้กำหนดโครงสร้างและไม่ได้กำหนดข้อสอบไว้ล่วงหน้า ว่าถ้าผู้สอบตอบถูกจะต้องไปทำข้อสอบข้อใดหรือถ้าตอบผิดจะต้องไปทำข้อสอบข้อใดแต่จะดำเนินการโดยใช้ข้อสอบที่กำหนดค่าฟังก์ชันสารสนเทศของข้อสอบไว้แล้ว การคัดเลือกข้อสอบเพื่อให้อันเหมาะสมกับความสามารถของผู้สอบจะดำเนินการโดยใช้วิธีการทางสถิติในการประมาณค่าความสามารถซึ่งวิธีการประมาณค่าความสามารถที่นิยมใช้มี 2 วิธี ได้แก่ การประมาณค่าแบบเบย์ (Bayesian Estimation) และการประมาณค่าแบบความเป็นไปได้สูงสุด (Maximum Likelihood Estimation)

การให้คะแนนการทดสอบแบบปรับเหมาะ

Weiss (1974 อ้างถึงใน สุพัฒน์ สุขมลสันต์, 2536) ได้กล่าวถึงวิธีการให้คะแนนการทดสอบแบบปรับเหมาะไว้หลายวิธี สรุปได้ดังนี้

1. ให้คะแนนตามค่าความยากของข้อสอบที่ยากที่สุดที่ตอบถูก (The most difficult of items answered correctly) เป็นการให้คะแนนโดยดูจากข้อสอบทั้งหมดที่ผู้สอบสามารถตอบได้ถูกต้อง หากข้อใดที่มีค่าความยากสูงสุดที่ผู้สอบตอบถูก ถือว่าเป็นค่าความสามารถของผู้สอบคนนั้น ค่าความยากดังกล่าวควรเป็นค่าความยากที่ได้จากการวิเคราะห์ข้อทดสอบตามทฤษฎีการตอบสนองข้อสอบ
2. ให้คะแนนตามค่าเฉลี่ยความยากของข้อที่ทำถูก (average difficulty of all items answered correctly) เป็นการให้คะแนนโดยการนำค่าความยากของข้อสอบทุกข้อที่ผู้สอบตอบถูกมาหาค่าเฉลี่ย ค่าเฉลี่ยความยากนี้ถือเป็นค่าความสามารถของผู้สอบคนนั้น
3. ให้คะแนนตามค่าเฉลี่ยความยากของทุกข้อที่ทำ (average difficulty) วิธีการให้คะแนนแบบนี้ Lord เป็นผู้เสนอ ในปี ค.ศ. 1970 โดยการนำค่าความยากของข้อสอบทุกข้อที่ผู้สอบได้ตอบโดยไม่คำนึงว่าจะตอบถูกหรือไม่มาหาค่าเฉลี่ย ค่าเฉลี่ยนี้ถือเป็นค่าความสามารถของผู้สอบ
4. ให้คะแนนตามค่าความยากของข้อสุดท้ายที่ตอบ (difficulty of the final item) นั่นคือไม่คำนึงว่าขั้นสุดท้ายจะตอบถูกหรือไม่ก็ตาม จะถือว่าคุณค่าความยากของข้อสอบข้อสุดท้ายที่ผู้สอบตอบคือค่าความสามารถของผู้สอบผู้นั้น
5. ให้คะแนนตามค่าความยากของข้อสอบที่ต่อจากขั้นสุดท้ายที่ตอบ (difficulty of $(N+1)^{th}$ item) ซึ่งมีลักษณะคล้ายกับวิธีที่ 4 แต่เป็นการตรวจสอบความสามารถต่อไปอีกขั้นหนึ่ง กล่าวคือ ถ้าขั้นสุดท้ายที่ผู้สอบสามารถตอบถูก คะแนนในขั้นต่อไปก็จะเพิ่มขึ้น แต่ถ้าขั้นสุดท้ายตอบผิดคะแนนขั้นต่อไปก็จะลดลง
6. ให้คะแนนตามกลวิธีของเบย์ (Bayesian strategies) ซึ่งมีหลายวิธีแต่ที่นิยมใช้กันมากคือ วิธีที่เรียกว่า “วิธีของเบย์ที่ปรับปรุงให้ทันสมัย” (Bayesian updating) ซึ่งสามารถประมาณค่าความสามารถของผู้สอบได้ค่อนข้างคงที่
7. ให้คะแนนโดยวิธีการประมาณค่าความเป็นไปได้สูงสุด (maximum likelihood) ซึ่งมีหลายวิธีแต่ที่นิยมใช้กันมากคือ วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข (conditional maximum likelihood) การ

ประมาณค่าวิธีนี้มีข้อจำกัดกรณีที่ผู้สอบตอบข้อสอบถูกหมดหรือผิดหมดจะไม่สามารถประมาณค่าได้
ขั้นตอนการให้คะแนนมี ดังนี้ (Hambleton and Swaminathan, 1985 อ้างถึงใน สุพัฒน์ สุกมลสันต์)

ขั้นที่ 1 ประมาณค่าความสามารถเริ่มต้น ($\theta_m; m = 0$) โดยใช้สูตร ดังนี้

$$\theta_0 = \ln[r_a / (k - r_a)]$$

เมื่อ $r_a = \sum a_i U_i;$ ($U_i = 1$) เมื่อตอบถูก

($U_i = 0$) เมื่อตอบผิด

a_i คือ อำนาจจำแนกของข้อสอบข้อที่ i

K คือ จำนวนข้อทดสอบทั้งหมดที่ผู้สอบตอบ

ขั้นที่ 2 หาค่า $P_i(\theta_m)$ และ $Q_i(\theta_m)$ โดยใช้สูตร ดังนี้

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

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

เมื่อ θ_m คือ ความสามารถของผู้สอบที่ประมาณค่าได้ในครั้งที่ m

a_i คือ ค่าอำนาจจำแนกของข้อทดสอบข้อที่ i

b_i คือ ค่าความยากของข้อสอบข้อที่ i

c_i คือ ค่าการเดาของข้อสอบข้อที่ i

D คือ ค่าคงที่ ซึ่งมีค่าเท่ากับ 1.7

e คือ ค่าคงที่ ซึ่งมีค่าเท่ากับ 2.7182

ขั้นที่ 3 หาค่าปรับแก้ (h_m) โดยใช้สูตร

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

ขั้นที่ 4 ประเมินค่าความสามารถของผู้สอบใหม่ ($\theta_m + 1$) โดยใช้สูตร

$$\theta_m + 1 = \theta_m - h_m$$

ขั้นที่ 5 คำนวณซ้ำในขั้นที่ 2, 3 และ 4 จนกระทั่ง h_m เข้าใกล้ศูนย์ ($h_m < 0.001$)

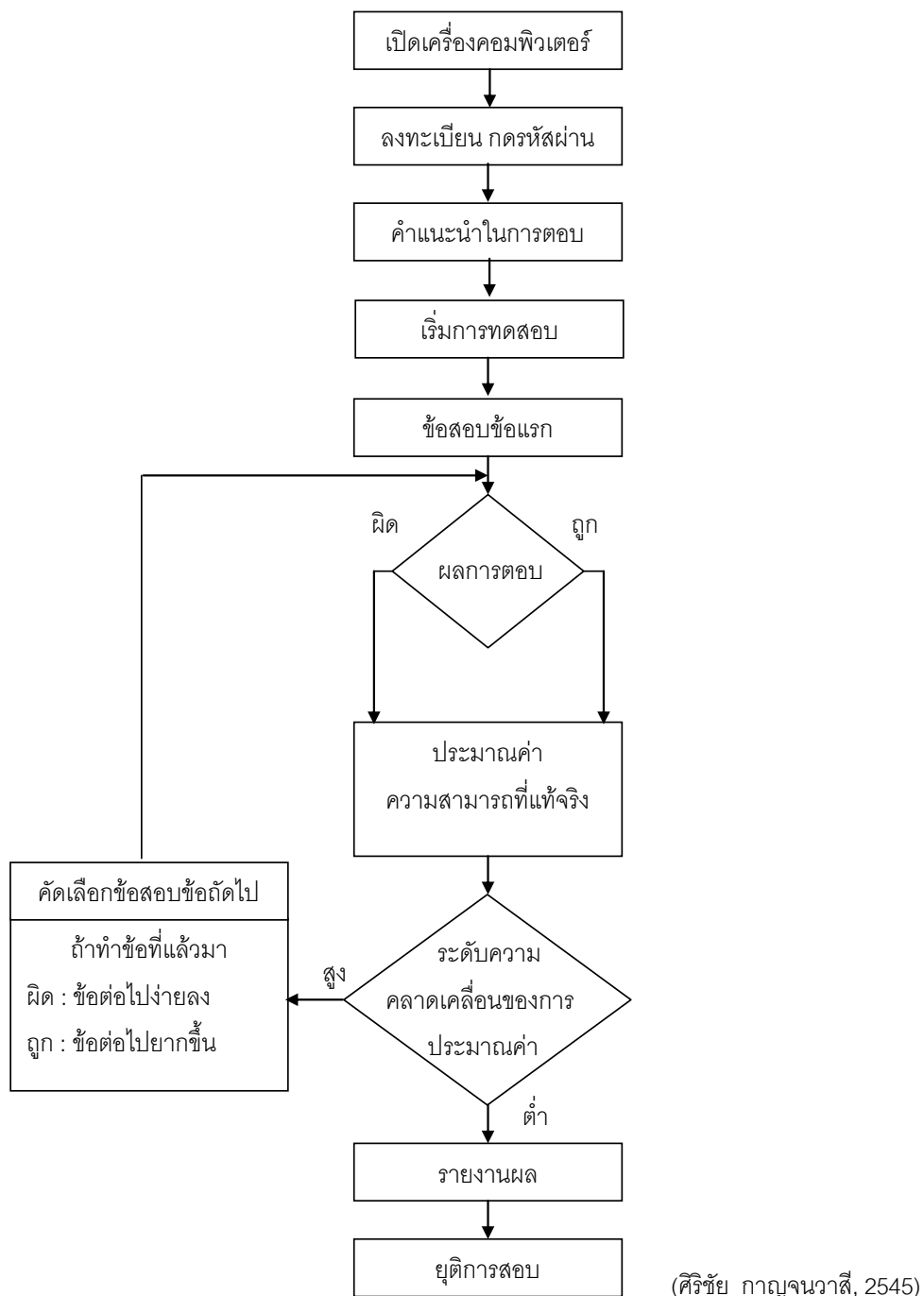
การประยุกต์ใช้คอมพิวเตอร์ในการทดสอบแบบปรับเหมาะ

การนำคอมพิวเตอร์มาใช้ในการทดสอบ (Computer Assisted Testing) เพื่อเพิ่มประสิทธิภาพในการตรวจข้อสอบ การวิเคราะห์ข้อสอบ การตัดเกรด การรายงานผลการสอบและการสร้างคลังข้อสอบมีการดำเนินการอย่างแพร่หลายระหว่างปี ค.ศ. 1970 ถึง ปี ค.ศ. 1980 (ศิริชัย กาญจนวาสี, 2545) สำหรับการทดสอบแบบปรับเหมาะเริ่มมีการประยุกต์คอมพิวเตอร์มาใช้ในการวิจัยการทดสอบแบบปรับเหมาะโดยได้รับการสนับสนุนจาก U.S. Office of Naval Research ตั้งแต่ปลายปี ค.ศ. 1960 ถึงต้นปี ค.ศ. 1970 (Weiss, 1983, 2004)

คอมพิวเตอร์เป็นเทคโนโลยีที่มีหน่วยความจำขนาดใหญ่สามารถเก็บรวบรวมข้อมูลทั้งตัวหนังสือและรูปภาพได้จำนวนมาก สามารถประมวลผลได้อย่างรวดเร็ว รวมทั้งมีปฏิสัมพันธ์กับผู้สอบได้อย่างมีประสิทธิภาพ คอมพิวเตอร์จึงมีความเหมาะสมและประโยชน์ยิ่งต่อการนำมาประยุกต์ใช้ในทดสอบแบบปรับเหมาะที่ต้องมีระบบการส่งผ่านซึ่งสามารถนำข้อสอบออกจากคลังข้อสอบและนำเสนอให้ผู้สอบตอบได้อย่างรวดเร็ว โดยต้องมีหลักเกณฑ์ในการคัดเลือกข้อสอบเริ่มต้น การประมาณค่าความสามารถของผู้สอบเพื่อคัดเลือกข้อสอบระหว่างกระบวนการทดสอบและการยุติการสอบที่ชัดเจน (van der Linden, 1999)

การประยุกต์ใช้คอมพิวเตอร์ในการทดสอบแบบปรับเหมาะเกิดขึ้นในทุกกระบวนการของการทดสอบตั้งแต่การสร้างคลังข้อสอบ การวิเคราะห์สารสนเทศของข้อสอบ การคัดเลือกข้อสอบ การประมาณค่าความสามารถของผู้สอบ การยุติการสอบ ข้อสอบจะถูกเก็บไว้ในคอมพิวเตอร์และแสดงออกทางหน้าจอ (monitor) เพื่อให้ผู้สอบตอบข้อสอบโดยผ่านทางแป้นพิมพ์ (keyboard) หรือ เมาส์ (mouse) โปรแกรมคอมพิวเตอร์จะทำหน้าที่เสมือนนักจิตวิทยาที่ดำเนินการสอบโดยใช้แบบทดสอบเชอร์ปัญญาของบีเนท์ซึ่งจะช่วยกำหนดข้อสอบเริ่มต้น ข้อสอบที่เหมาะสมกับความสามารถของผู้สอบในระหว่างกระบวนการทดสอบ และตัดสินใจยุติการสอบ (Weiss, 2004)

ขั้นตอนในการดำเนินการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์จะเริ่มจากการลงทะเบียนสอบและการกรอกรหัสผ่านบนแป้นพิมพ์ของคอมพิวเตอร์ ก่อนเริ่มการทดสอบคอมพิวเตอร์จะแนะนำวิธีการตอบข้อสอบ มีการชั่งชั่งและทดลองการตอบ เมื่อผู้สอบมีความเข้าใจดีแล้วการสอบก็จะเริ่มขึ้น โดยการสุ่มข้อสอบข้อแรก (initial item) จากคลังข้อสอบซึ่งตามหลักการทั่วไปจะเป็นข้อสอบที่มีความยากปานกลางหรือเป็นข้อสอบที่คัดเลือกให้เหมาะสมกับผู้สอบตามสารสนเทศที่ได้รับ หลังจากตรวจข้อสอบจะมีการประมาณค่าความสามารถเบื้องต้นของผู้สอบ ถ้าตอบถูกข้อถัดไปจะยากขึ้น แต่ถ้าตอบผิดข้อถัดไปจะง่ายลง กระบวนการทดสอบจะดำเนินการเช่นนี้ซ้ำไปเรื่อย ๆ จนกว่าจะถึงเกณฑ์การยุติการสอบที่กำหนดไว้ โปรแกรมจะยุติการทำงาน บันทึกผลการสอบ รายงานผลความสามารถที่แท้จริงของผู้สอบบนหน้าจอคอมพิวเตอร์และพิมพ์รายงานไปยังผู้ที่เกี่ยวข้อง (ศิริชัย กาญจนวาสี, 2538, 2545) ดังแสดงในแผนภูมิที่ 3



แผนภูมิที่ 3 ขั้นตอนการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์

องค์ประกอบของการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์

องค์ประกอบของการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ มีรายละเอียด ดังนี้

1. จุดเริ่มต้นการทดสอบ (starting point)

การทดสอบควรเริ่มต้นด้วยข้อสอบที่ตรงกับความสามารถของผู้สอบหรือใกล้เคียงกับความสามารถของผู้สอบมากที่สุด ซึ่งจะช่วยให้จำนวนข้อสอบที่ใช้ในการทดสอบสำหรับแต่ละบุคคลมีจำนวนลดลง (นันทิยา พึ่งคำ, 2531) การเลือกข้อสอบข้อแรกอาจพิจารณาจากผลสัมฤทธิ์ที่ผ่านมาของผู้สอบ (Hambleton, zaal และ Pieters (2000 อ้างถึงใน Latu and Chapman, 2002) โดยทั่วไปนิยมเลือกข้อสอบที่มีความยากปานกลาง โดยอาจดำเนินการได้ใน 2 กรณี กรณีแรก กลุ่มผู้สอบมีความสามารถใกล้เคียงกัน (homogeneous) หรือไม่มีข้อมูลสำหรับผลสัมฤทธิ์ที่ผ่านมาควรเริ่มต้นด้วยข้อสอบที่มีค่าความยากปานกลางเท่ากันสำหรับทุกคน กรณีที่สอง ผู้สอบมีความสามารถแตกต่างกัน (heterogeneous) หรือมีข้อมูลเกี่ยวกับการศึกษาของผู้สอบบ้าง ควรใช้ค่าความยากปานกลางที่เหมาะสมสำหรับผู้สอบแต่ละกลุ่มความสามารถ โดยพิจารณาจากข้อมูลที่มีอยู่ เช่น กลุ่มอายุ ระดับชั้นที่ศึกษา เป็นต้น (ศิริชัย กาญจนวาสี, 2545)

ความสำคัญของจุดเริ่มต้นยังคงเป็นที่โต้แย้งกันในหมู่นักวิจัย บางคนกล่าวว่าจุดเริ่มต้นไม่มีผลกระทบต่อการสอบถ้ามีจำนวนข้อสอบมากพอ เช่น Lord (1977) อ้างถึงใน ศิริชัย, 2545) ได้ทำการทดลองด้วยสถานการณ์จำลอง โดยเริ่มต้นด้วยข้อสอบที่มีค่าความยากระดับต่าง ๆ พบว่าเกือบไม่มีผลต่อความถูกต้องแม่นยำในการประมาณค่าความสามารถของผู้สอบเลย โดยเฉพาะกรณีที่มีข้อสอบในคลังข้อสอบจำนวนมาก แต่บางคนกล่าวว่าจุดเริ่มต้นที่ไม่เหมาะสมจะทำให้ผู้สอบเกิดความวิตกกังวลและรู้สึกผิดหวังซึ่งมีผลกระทบต่อการทำแบบทดสอบ (Hambleton, zaal และ Pieters 2000 อ้างถึงใน Latu and Chapman, 2002)

2. การคัดเลือกข้อสอบและการประมาณค่าความสามารถ (item selection and ability estimate)

การคัดเลือกข้อสอบและการประมาณค่าความสามารถของผู้สอบมีประสิทธิภาพมากขึ้นเมื่อนำทฤษฎีการตอบสนองข้อสอบ (item response theory) มาประยุกต์ใช้ โดยทั่วไปนิยมใช้วิธีที่สอดคล้องกันเพื่อความสะดวกรวดเร็วในการคำนวณ วิธีที่นิยมนำมาใช้ในการประมาณค่าความสามารถมี 2 วิธี ได้แก่ วิธีการประมาณค่าความเป็นไปได้สูงสุด (Maximum Likelihood Estimate) และการประมาณค่าแบบเบย์ (Bayesian Estimate) วิธีการคัดเลือกข้อสอบที่ใช้สำหรับการประมาณค่าความเป็นไปได้สูงสุด ประกอบด้วยวิธีการคัดเลือกข้อสอบที่มีค่าความยากสอดคล้องกับค่าความสามารถที่ประมาณได้ การคัดเลือกข้อสอบที่มีตำแหน่งสารสนเทศ (ค่าความยาก ค่าอำนาจจำแนกและค่าการเดา) สูงสุดสอดคล้องกับความสามารถที่ประมาณได้และการคัดเลือกข้อสอบที่ให้สารสนเทศสูงสุดที่ตำแหน่งความสามารถนั้น ส่วนวิธีการคัดเลือกข้อสอบที่ใช้สำหรับการประมาณค่าแบบเบย์ใช้การคัดเลือกข้อสอบข้อที่ยังไม่ได้นำมาใช้ในการทดสอบนั้นที่จะให้ค่าความแปรปรวนของค่าความสามารถที่คาดหวังมีค่าต่ำสุด (นันทิยา พึ่งคำ, 2531; สายชล อบทม, 2539; ศิริชัย กาญจนวาสี, 2545)

3. เกณฑ์การยุติการสอบ (termination criteria or Stopping criteria)

ปัจจุบันมีการพัฒนาวิธีที่ใช้ในการยุติการสอบหลายวิธี แต่วิธีที่นิยมใช้มี 2 วิธี วิธีแรก คือ การทดสอบจนกระทั่งได้ระดับความคงที่ในการวัด (level of measurement stability or consistency) โดยการกำหนดระดับความคลาดเคลื่อนที่ยอมรับได้ วิธีที่สอง คือ การกำหนดจำนวนข้อสอบให้คงที่ วิธีการกำหนดความยาวของแบบสอบให้คงที่ช่วยให้แน่ใจได้ว่าผู้สอบได้ทำข้อสอบในจำนวนข้อสอบที่เท่ากันและช่วยให้แบบทดสอบไม่ยาวเกินไป แต่ให้คุณภาพของการวัดผลที่แตกต่างกัน เนื่องจากมีระดับความยากของข้อสอบแตกต่างกัน การใช้การยุติการสอบโดยพิจารณาจากค่าความคลาดเคลื่อนในการวัดจึงเป็นวิธีที่ดีกว่า (ศิริชัย กาญจนวาสี, 2545; Hambleton, zaal และ Pieters 2000 อ้างถึงใน Latu and Chapman, 2002)

4. จำนวนและเนื้อหาของข้อสอบในคลังข้อสอบ (Size and content of item banks)

ความตรงของแบบทดสอบขึ้นอยู่กับความเกี่ยวข้องและความเหมาะสมของข้อสอบที่ใช้ แต่สำหรับการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ความตรงของแบบทดสอบจะขึ้นอยู่กับทั้งจำนวนและเนื้อหาของข้อสอบในคลังข้อสอบ นอกจากข้อสอบจะต้องมีจำนวนมากเพียงพอในการประมาณค่าระดับความสามารถของผู้สอบซึ่งมีความแตกต่างกันได้อย่างถูกต้องเหมาะสมแล้ว ข้อควรคำนึงที่สำคัญประการหนึ่ง ก็คือ เนื้อหาของข้อสอบ เนื่องจากการมีข้อสอบในคลังจำนวนมากทำให้มีโอกาสที่จะมีข้อสอบที่ไม่ดีได้มาก ซึ่งข้อสอบเหล่านั้นจะมีผลกระทบอย่างมากต่อการให้คะแนนสุดท้ายในการทดสอบแบบปรับเหมาะ เพราะในการทดสอบแบบปรับเหมาะมีการตรวจให้คะแนนในทันที ข้อสอบที่ถูกเลือกให้ผู้สอบทำจึงมีผลต่อความถูกต้องในการประมาณค่าความสามารถของผู้สอบและการให้คะแนนสุดท้ายโดยไม่มีการแก้ไขหรือย้อนกลับได้ (Hambleton, zaal และ Pieters 2000 อ้างถึงใน Latu and Chapman, 2002)

ประโยชน์ของการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์

Green (1983 อ้างถึงใน Wainer, 1990) และ Latu และ Chapman (2002) ได้กล่าวถึงประโยชน์ของการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ไว้ สรุปได้ดังนี้

1. ช่วยเพิ่มประสิทธิภาพในการทดสอบ โดยเฉพาะภาระงานการทดสอบต่าง ๆ ได้แก่ การดำเนินการทดสอบ การกำหนดเกณฑ์และการให้คะแนน
2. แบบทดสอบที่เก็บไว้ในคลังข้อสอบในคอมพิวเตอร์จะมีความปลอดภัย (security) ต่อการคัดลอกมากกว่า
3. ไม่มีข้อจำกัดเรื่องเวลา ผู้สอบสามารถทำแบบทดสอบในเวลาที่แตกต่างกันตามความพร้อมของแต่ละบุคคลภายในช่วงเวลาที่กำหนด

4. ผู้สอบไม่เกิดความท้อแท้ในการทำข้อสอบ เนื่องจากเป็นการคัดเลือกข้อสอบที่เหมาะสมกับความสามารถของแต่ละบุคคล
5. การตรวจข้อสอบมีความชัดเจนขึ้น เนื่องจากไม่มีปัญหาที่เกิดจากกระดาษคำตอบ เช่น การทำเครื่องหมายไม่ชัดเจน ความไม่ชัดเจนในข้อที่เลือกเนื่องจากการลบเมื่อต้องการเปลี่ยนคำตอบ เป็นต้น
6. ข้อสอบสามารถตรวจให้คะแนนและให้ผลย้อนกลับ (feedback) กับผู้สอบได้ทันที
7. สามารถคัดเลือกข้อสอบจากการตอบข้อสอบเบื้องต้นได้อย่างรวดเร็ว
8. ผู้สอบได้ทำข้อสอบที่มีความยาวเหมาะสมกับระดับความสามารถ
9. มีการเสนอข้อสอบให้ผู้สอบตอบในรูปแบบที่หลากหลายทั้งภาพนิ่ง ภาพเคลื่อนไหวและเสียงประกอบ

ข้อจำกัดของการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์

Latu และ Chapman (2002) กล่าวถึงข้อจำกัดของการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ไว้ดังนี้

1. ความปลอดภัยของแบบทดสอบ (Test security) แม้ว่าจะมีนักวิจัยหลาย ๆ คน แสดงให้เห็นว่าการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์มีความปลอดภัยของแบบทดสอบมากกว่าการทำแบบทดสอบบนโต๊ะ (desk drawer) แต่ก็มีนักวิจัยบางท่านโต้แย้งว่าปัจจุบันมีวิธีการขโมยข้อสอบได้หลายวิธี เช่น pager กล้องขนาดเล็ก (miniature cameras) กล้องวิดีโอ (video transmitters) เครื่องบันทึกวิดีโอขนาดเล็ก (micro video recorders) นอกจากนี้ผู้สอบสามารถคัดลอกข้อสอบจากเครื่องคอมพิวเตอร์ได้ง่ายและรวดเร็วกว่าการคัดลอกข้อสอบจากกระดาษคำตอบอีกด้วย
2. ปัญหาเกี่ยวกับผู้สอบ (Examinee issues) แม้ว่าการทดสอบแบบปรับเหมาะจะมีประโยชน์ต่อผู้สอบมาก แต่ยังคงมีข้อจำกัดเกี่ยวกับผู้สอบ เช่น ผู้สอบหลายคนไม่เคยชินกับการใช้คอมพิวเตอร์ ทำให้เกิดความวิตกกังวล ผู้สอบบางคนแสดงให้เห็นว่าการนำเสนอโดยใช้คอมพิวเตอร์เป็นสิ่งที่ยุ่งยาก (difficult or fatiguing) โดยเฉพาะการตอบโดยใช้เมาส์หรือคีย์บอร์ด
3. ค่าใช้จ่ายในการดำเนินการ (Financial realities) การทดสอบแบบปรับเหมาะมีค่าใช้จ่ายในการดำเนินการสร้างแบบทดสอบการเตรียมการและการดำเนินการสอบที่ค่อนข้างสูง เช่น Educational Testing Service (ETS) ศึกษาค่าใช้จ่ายในการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์พบว่าค่าใช้จ่ายที่ต้องจ่ายจริงมากกว่าวิธีการทดสอบแบบประเพณีนิยม ดังจะเห็นได้จากในสหรัฐอเมริกาผู้สอบจะต้องเสียค่าใช้จ่ายมากกว่า 100 ดอลลาร์ เพื่อสอบ TOFEL ในระบบการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ ค่าใช้จ่ายที่แพงมากดังกล่าวมาจากการพัฒนาคลังข้อสอบ การบริหารการทดสอบและการพัฒนาโปรแกรมสำเร็จรูป ซึ่งมีค่าใช้จ่ายในการพัฒนาและการนำไปใช้ในระบบที่สมบูรณ์ระหว่าง 3,000 -5,000 ดอลลาร์

รายการอ้างอิง

- นันทิยา พึ่งคำ. (2531). การเปรียบเทียบคุณภาพการทดสอบแบบ ซี เอ ที และแบบประเพณีนิยม ในการวัดความสามารถด้านคำศัพท์ภาษาอังกฤษของนักเรียนชั้นมัธยมศึกษาปีที่ 3. วิทยานิพนธ์ปริญญาโทมหาบัณฑิต. ภาควิชาวิจัยการศึกษา. บัณฑิตวิทยาลัย. จุฬาลงกรณ์มหาวิทยาลัย.
- ศิริชัย กาญจนวาสี. (2538). การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบ. กรุงเทพฯ: ศูนย์ตำราและเอกสารทางวิชาการ คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.
- ศิริชัย กาญจนวาสี. (2545). ทฤษฎีการทดสอบแนวใหม่. พิมพ์ครั้งที่ 2. กรุงเทพฯ: สายชล อบทม. (2539). การพัฒนาโปรแกรมสำหรับการทดสอบแบบปรับเหมาะโดยใช้ คอมพิวเตอร์. วิทยานิพนธ์ปริญญาโทมหาบัณฑิต. ภาควิชาวิจัยการศึกษา. บัณฑิตวิทยาลัย. จุฬาลงกรณ์มหาวิทยาลัย. โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- Weiss,D.J. (1983). Introduction. In Weiss,D.J. (eds.), New horizons in testing : latent trait test theory and computerized adaptive testing, pp. 1-7. New York: ACADEMIC.
- Weiss,D.J. (2004). Computerized adaptive testing for effective and efficient measurement in counseling and education. Measurement and Evaluation in Counseling and Development, 37:70-84.
- Weiss, D.J. and Schleisman, J.L. (1999). Adaptive testing. In Masters, G.N. and Keeves, J.P. (eds.), Advances in Measurement in Educational Research and Measurement, pp. 138-150. Amsterdam: PERGAMON