

การพัฒนาการทดสอบเฉพาะบุคคลแบบ ซี เอ ที

Development of Tailored Testing with Computerized Adaptive Testing

สุรศักดิ์ อมรรัตนศักดิ์¹



บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อพัฒนาการทดสอบเฉพาะบุคคลแบบ ซี เอ ที โดยใช้กลยุทธ์หลายขั้นตอน รูปแบบปิรามิด แบบมีหลายข้อในแต่ละชั้นและหาคุณภาพของแบบสอบที่สร้างขึ้น กลุ่มตัวอย่างเป็นนักศึกษา มหาวิทยาลัยรามคำแหงชั้นปีที่ 1 จำนวน 2,868 คน ในการวิจัยครั้งนี้มี 4 ระยะ คือ ระยะที่หนึ่งเป็นการทดลองกลุ่มข้อสอบ ระยะที่สองเป็นการหาค่าพารามิเตอร์ของข้อสอบตามทฤษฎีการตอบสนองต่อข้อสอบแบบ 3 พารามิเตอร์ ระยะที่สามเป็นการทดลองใช้แบบสอบเฉพาะบุคคลโดยใช้เครื่องคอมพิวเตอร์ในการทดสอบ และระยะที่สี่เป็นการหาคุณภาพของแบบสอบเฉพาะบุคคล ผลการวิจัย พบว่า 1. แบบสอบเฉพาะบุคคลแบบ ซี เอ ที วิชา EN101 มี 152 ข้อ โดยจัดกลุ่มข้อสอบตามค่าความยากออกเป็น 11 ระดับ ๆ ละ 6-43 ข้อ มีค่าอำนาจจำแนกอยู่ระหว่าง 0.500 ถึง 2.681 ค่าความยากอยู่ระหว่าง -3.614 ถึง 3.997 และค่าการเดาอยู่ระหว่าง 0.053 ถึง 0.300 2. สัมประสิทธิ์สหสัมพันธ์ระหว่างค่าความสามารถที่ประมาณได้จากแบบสอบเฉพาะบุคคลแบบ ซี เอ ที กับคะแนนจากแบบสอบแบบคลาสสิก มีค่าเท่ากับ 0.813 3. แบบสอบเฉพาะบุคคลแบบ ซี เอ ที มีความเที่ยงตรงเชิงสภาพโดยมีค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนความสามารถที่ประมาณได้จากแบบสอบเฉพาะบุคคลกับเกรดวิชา EN101 เท่ากับ 0.772

คำสำคัญ: ซี เอ ที การพัฒนาแบบสอบเฉพาะบุคคล การทดสอบเฉพาะบุคคล

ABSTRACT

The purposes of this study were to develop a tailored test with a computerized adaptive test estimated by multi-stages strategies with multiple-item pyramidal model and to find the quality of the test. The samples were 2,868 first year students of Ramkhamhaeng University. There were 4 stages in developing test - to try out the items pool, to estimate item parameter by 3 parameters item response theory, to try out the computerized adaptive test, and to find the quality of the test. The findings of this research were as follows: 1. There were 152 items in EN101 computerized adaptive tailored test with 11 levels of difficulty and about 6-43 items in each level. The ranges of their discriminating power, item difficulty and guessing indices were from 0.500 to 2.681, from -3.614 to 3.997 and from 0.053 to 0.300, respectively. 2. The correlation coefficient between the ability estimated from tailored test and the score from classical test was 0.813. 3. The concurrent validity of the EN101 computerized adaptive tailored test determined from the correlation coefficient between the ability estimated from tailored test and EN101 grade was 0.772.

keywords: computerized adaptive testing, tailored test Development, tailored testing

¹ รองศาสตราจารย์ ดร. ภาควิชาการประเมินและการวิจัย คณะศึกษาศาสตร์ มหาวิทยาลัยรามคำแหง

บทนำ

ในการวัดผลทางการศึกษาและจิตวิทยานั้น สิ่งที่ต้องการจะวัดและตรวจสอบมักเป็นคุณลักษณะ (Traits) ต่างๆ ที่เป็นคุณลักษณะภายใน (Latent traits) ที่ไม่สามารถสังเกตได้โดยตรง เช่น ความสามารถในการคำนวณ ทักษะในการอ่านจับใจความ เป็นต้น ข้อสอบเป็นเพียงสิ่งเร้าที่กระตุ้นให้บุคคลแสดงพฤติกรรมออกมาโดยการตอบข้อสอบ (Response) เหล่านั้น แล้วก็พิจารณาว่าบุคคลนั้นทำข้อสอบได้ถูกต้องกี่ข้อ ได้กี่คะแนน ต่อจากนั้นก็อนุมาน (Infer) จากจำนวนข้อที่ถูกโยงไปหาความสามารถที่แท้จริง ซึ่งเป็นคุณลักษณะภายในของผู้สอบ (สงบ, 2525)

แบบสอบที่ใช้วัดคุณลักษณะภายในของผู้สอบเท่าที่ผ่านมามักจะยึดทฤษฎีการวัดแบบคลาสสิก (Classical test theory) ที่ประกอบด้วยกลุ่มข้อสอบที่มีข้อสอบจำนวนมากและจัดเรียงลำดับข้อสอบที่แน่นอน ผู้สอบทุกคนจะต้องทำข้อสอบทุกข้อในแบบสอบเหมือนกันหมด ซึ่งแบบสอบดังกล่าวนี้จะประกอบด้วยข้อสอบที่มีความยากพอเหมาะกับผู้สอบที่มีความสามารถปานกลางเท่านั้น (Vale and Weiss, 1975) ดังนั้น ในการทดสอบผลการวัดจะมีความถูกต้องแม่นยำ เมื่อผู้สอบมีความสามารถปานกลาง และเมื่อใดก็ตามที่ความสามารถของผู้สอบเบี่ยงเบนไปจากความสามารถปานกลาง ความคลาดเคลื่อนของการวัดก็จะเพิ่มขึ้น แบบสอบที่ยึดทฤษฎีการวัดแบบคลาสสิกนั้นมุ่งให้ผู้สอบทุกคนทำข้อสอบทุกข้อที่มีอยู่ในแบบสอบ โดยไม่คำนึงว่าข้อสอบเหล่านั้นจะเหมาะสมกับความสามารถของผู้สอบหรือไม่ หากพิจารณาถึงผลกระทบในเชิงจิตวิทยาจะพบว่าถ้าผู้สอบต้องสอบข้อสอบที่ยากเกินความสามารถแล้วจะทำให้ผู้สอบเกิดความท้อถอยในการทำข้อสอบและมีแนวโน้มที่จะตอบโดยการเดาสูง ส่วนข้อสอบที่ง่ายเกินไปก็ไม่ทำให้ผู้สอบได้ใช้ความสามารถอย่างเต็มที่ ทำให้ผู้สอบเกิดความเบื่อหน่าย ผลกระทบเหล่านี้ส่งผลให้ความถูกต้องแม่นยำของคะแนนลดต่ำลง และเป็นผลทำให้การประมาณค่าความสามารถของผู้สอบคลาดเคลื่อนไปจากความจริง (Weiss, 1974)

จะเห็นได้ว่าทฤษฎีการวัดแบบคลาสสิก ซึ่งเป็นทฤษฎีการวัดที่นิยมใช้วัดกันมาเป็นเวลานานนั้นมีพื้นฐานอยู่บนข้อตกลงเบื้องต้นที่ไม่แกร่งเท่าที่ควร (Weak assumption) ซึ่ง Hambleton และ Swaminathan (1985) ได้สรุปจุดอ่อนของทฤษฎีการวัดแบบคลาสสิกไว้ดังนี้

1. ค่าสถิติของข้อสอบ ได้แก่ ค่าความยากและค่าอำนาจจำแนกจะมีค่าแปรเปลี่ยนไปตามลักษณะของกลุ่มตัวอย่างที่นำมาใช้ในการวิเคราะห์ นั่นคือถ้ากลุ่มตัวอย่างเปลี่ยนแปลงไปค่าสถิติของข้อสอบก็จะมีค่าเปลี่ยนแปลงไป ดังนั้น ผลการวิเคราะห์จึงใช้ได้จำกัดเฉพาะกลุ่มผู้ที่มีลักษณะเหมือนกับกลุ่มตัวอย่างที่นำมาใช้ในการวิเคราะห์เท่านั้น

2. การวิเคราะห์คุณภาพของแบบสอบ หรือการประมาณค่าความสามารถของผู้สอบอยู่บนพื้นฐานของการใช้แบบสอบฉบับเดียวกันหรือแบบสอบที่คู่ขนานกัน (Parallel tests) ถ้าใช้แบบสอบต่างฉบับกันจะนำผลการวิเคราะห์หรือคะแนนสอบมาเปรียบเทียบกันไม่ได้ ดังนั้น การเปรียบเทียบความสามารถระหว่างผู้สอบจึงทำได้เฉพาะกลุ่มผู้สอบที่ใช้แบบสอบคู่ขนานหรือแบบสอบชุดเดียวกัน

3. ความแปรปรวนของความคลาดเคลื่อนในการวัด (Error variance) จะมีค่าคงที่ตลอดสำหรับผู้สอบทุกระดับความสามารถ ซึ่งอาจจะไม่จริงเพราะในแต่ละระดับความสามารถของผู้เรียนน่าจะมีระดับของความคลาดเคลื่อนเกิดขึ้นแตกต่างกันได้

4. ทฤษฎีการวัดแบบคลาสสิกไม่มีการศึกษาถึงความสัมพันธ์ระหว่างระดับความสามารถของผู้สอบกับพฤติกรรมคำตอบข้อสอบเป็นรายข้อว่ามีความน่าจะเป็นหรือโอกาสในการตอบข้อสอบถูกมีมากน้อยอย่างไร ไม่มีสารสนเทศพื้นฐานที่จะช่วยทำนายพฤติกรรมคำตอบสนองข้อสอบของผู้สอบ จึงมีประโยชน์จำกัดในการพัฒนาข้อสอบ การคัดเลือกข้อสอบให้เหมาะสมกับระดับความสามารถของผู้สอบ การเปรียบเทียบคะแนนระหว่างแบบสอบต่างฉบับ การตรวจสอบความลำเอียงของข้อสอบ เป็นต้น

จากจุดอ่อนของทฤษฎีการวัดแบบคลาสสิกที่กล่าวมาจึงทำให้ทฤษฎีนี้ถูกมองว่าเป็นทฤษฎีที่ไม่เพียงพอต่อการประมาณความสามารถที่แท้จริงและไม่สามารถแก้ปัญหาอีกหลายประการของการวัดผลในปัจจุบัน นักทฤษฎีทางการวัดผลจึงได้ช่วยกันคิดและสร้างทฤษฎีขึ้นมาใหม่เพื่อแก้ปัญหาดังกล่าว ทฤษฎีหนึ่งที่นักวัดผลช่วยกันพัฒนาขึ้นมาคือทฤษฎีการตอบสนองต่อข้อสอบ (Item response theory) ผู้ที่ทำการบุกเบิกทฤษฎีนี้กันอย่างจริงจังก็คือ Lord, Birbaum และ Rasch ทฤษฎีการตอบสนองต่อข้อสอบนี้บางทีก็เรียกทฤษฎีความสามารถแฝง (Latent trait theory) หรือทฤษฎีโค้งลักษณะข้อสอบ (Item characteristic curve theory)

ตามที่ได้กล่าวมาแล้วว่าแบบสอบที่ยึดทฤษฎีการวัดแบบคลาสสิกนั้น จะใช้ได้เหมาะกับผู้ที่มีความสามารถในระดับปานกลางเท่านั้น แต่หากผู้สอบที่มีความสามารถสูงหรือต่ำมาก ๆ แล้วแบบสอบที่ยึดทฤษฎีการวัดแบบคลาสสิกก็จะมีผลทำให้ความเชื่อมั่นของคะแนนผลการสอบต่ำกว่าที่ควรจะเป็น ปัญหาเรื่องนี้สามารถแก้ไขโดยการประยุกต์ใช้ทฤษฎีการตอบสนองต่อข้อสอบ

การทดสอบเฉพาะบุคคล เป็นการทดสอบที่พยายามคัดเลือกข้อสอบที่มีระดับความยากเหมาะสมกับระดับความสามารถของผู้สอบ การสอบครั้งหนึ่งผู้สอบแต่ละคนไม่จำเป็นต้องสอบข้อสอบเหมือนกันทุกข้อทั้งฉบับ ผู้สอบแต่ละคนจะตอบข้อสอบเพียงบางข้อโดยจะตอบในข้อที่เหมาะสมกับระดับความสามารถ ทั้งนี้เนื่องจากยึดหลักที่ว่าถ้าผู้สอบตอบข้อสอบที่ยากได้ถูกต้อง ก็น่าที่จะตอบข้อสอบข้อที่ง่ายกว่านั้นได้ จึงไม่จำเป็นต้องทำข้อสอบที่ง่ายกว่า ในทางตรงกันข้ามหากผู้สอบตอบข้อสอบง่าย ๆ ผิดแล้วข้อสอบข้อที่ยากกว่านั้นก็ไม่น่าจะผิดด้วย จึงไม่จำเป็นต้องทำข้อที่ยาก ๆ ดังนั้นในการทดสอบเฉพาะบุคคลจึงต้องมีการประมาณระดับความสามารถของผู้สอบก่อน วิธีการที่ง่ายและสะดวกก็คือให้ผู้สอบเริ่มทำข้อสอบที่มีความยากปานกลางก่อน ถ้าตอบผิดก็ให้ตอบข้อสอบข้อที่ง่ายกว่านั้นเป็นข้อต่อไป แต่ถ้าตอบถูกก็ให้ตอบข้อที่ยากกว่านั้น

ดำเนินการเช่นนี้จนสิ้นสุดการสอบ (Hambleton and Swaminathan, 1985; Weiss and Kingsbury, 1984)

การทดสอบเฉพาะบุคคลแตกต่างจากวิธีการทดสอบแบบคลาสสิก เนื่องจากการทดสอบเฉพาะบุคคลจะต้องกำหนดส่วนประกอบต่าง ๆ ที่ใช้ในการดำเนินการสอบไว้ล่วงหน้าทั้งสิ้น ส่วนประกอบเหล่านี้ได้แก่ รูปแบบการตอบข้อสอบ กลุ่มข้อสอบ ระดับในการเริ่มต้นทำข้อสอบ กฎในการคัดเลือกข้อสอบ วิธีการให้คะแนน และเกณฑ์ในการยุติการสอบ เนื่องจากความพยายามที่จะหาวิธีการและรูปแบบที่มีประสิทธิภาพในการทดสอบแนวใหม่จึงต้องมีกลยุทธ์ต่าง ๆ ในการสร้างและดำเนินการสอบ

จากการศึกษาของ Betz และ Weiss (1973) พบว่าการทดสอบเฉพาะบุคคลจะใช้ข้อสอบน้อยกว่าการทดสอบแบบคลาสสิก จึงใช้เวลาในการทำแบบสอบน้อยลง ในขณะที่ความเชื่อมั่นและความเที่ยงตรงคงเดิม ทั้งนี้เพราะผู้สอบไม่เกิดความเบื่อหน่าย ช่วยให้ผู้สอบเกิดความตั้งใจที่จะทำข้อสอบ มีแรงจูงใจในการสอบสูงขึ้น โดยใช้ความสามารถที่แท้จริงอย่างเต็มที่อยู่ตลอดเวลา เป็นผลให้ได้คะแนนที่ได้เกิดจากความสามารถที่แท้จริง

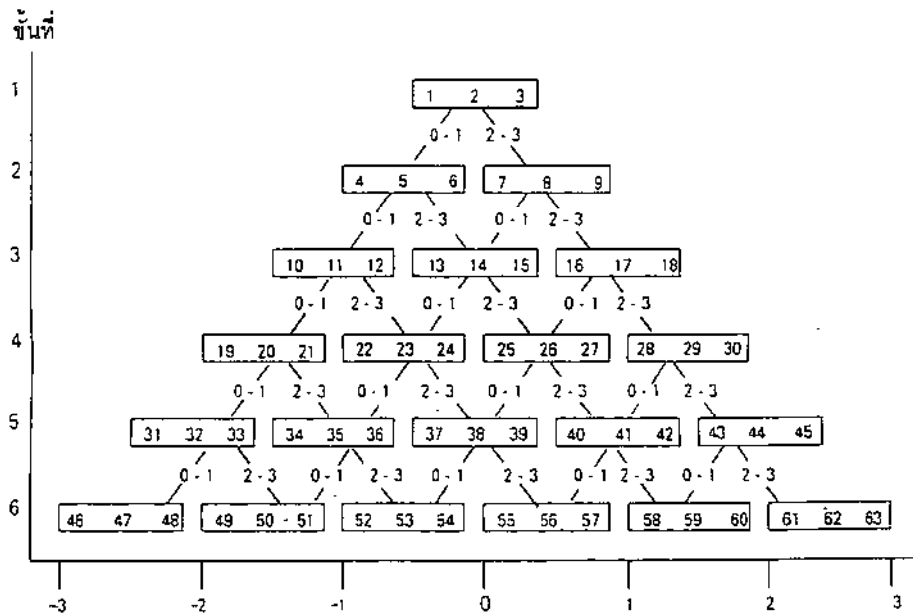
บุญชม (2535) ได้ตั้งข้อสังเกตเกี่ยวกับการทดสอบแบบเทอร์หรือการทดสอบเฉพาะบุคคลว่าการนำทฤษฎีการตอบสนองต่อข้อสอบมาใช้ในการประมาณค่าความสามารถของผู้สอบ ทำให้ประมาณค่าความสามารถของผู้สอบได้เที่ยงตรงและเชื่อถือได้ยิ่งขึ้น ส่วนการสร้างแบบสอบและกระดาษ คำตอบ มีวิธีการยากกว่าแบบสอบแบบเดิมที่ใช้โดยทั่วไป การตอบก็แตกต่างไปจากแบบเดิมที่ผู้สอบเคยมีประสบการณ์ จึงยากต่อการนำไปใช้จริง เว้นแต่จะใช้ทดสอบโดยใช้ระบบคอมพิวเตอร์

การที่จะก้าวให้ทันโลกยุคปัจจุบันจำเป็นต้องพัฒนาการทดสอบให้คู่ไปกับการพัฒนาสื่อการสอนที่ได้มีการนำเทคโนโลยีใหม่ ๆ เข้ามาใช้ในการเรียนการสอน ดังนั้น การนำเอาคอมพิวเตอร์มาใช้ในการสอบจึงเป็นสิ่งที่เหมาะสมในวงการศึกษายุคโลกาภิวัตน์จากการศึกษารูปแบบการทดสอบเฉพาะบุคคล พบว่า

เป็นเรื่องที่น่าสนใจและงานวิจัยเกี่ยวกับเรื่องนี้มีน้อย ส่วนมากจะเป็นการทดสอบในกระดาษเขียนตอบซึ่งมีปัญหาในการดำเนินการสอบ ประกอบกับในปัจจุบันมีความสะดวกในการวิเคราะห์ค่าพารามิเตอร์ต่าง ๆ มากขึ้น และสิ่งสำคัญ คือ การพัฒนาด้านคอมพิวเตอร์และการเขียนโปรแกรมคอมพิวเตอร์ที่ทำให้สามารถดำเนินการทดสอบตามวิธีการของการทดสอบเฉพาะบุคคลแบบ ซี เอ ที (Computerized adaptive testing) เพื่อเป็นแนวทางในการใช้คอมพิวเตอร์ในการทดสอบที่สามารถนำไปใช้ได้จริง โดยเป็นการนำข้อสอบจากคลังข้อสอบมาใช้ให้เหมาะสมกับความสามารถของผู้สอบและเป็นการลดจำนวนข้อสอบและเวลาในการสอบลง

วิธีดำเนินการวิจัย

การวิจัยเรื่องนี้มีวัตถุประสงค์เพื่อพัฒนาการทดสอบเฉพาะบุคคลแบบ ซี เอ ที ในวิชา EN101 ของมหาวิทยาลัยรามคำแหงโดยในการวิจัยครั้งนี้ผู้วิจัยจะใช้กลยุทธ์หลายขั้นตอน (Multi-stages strategies รูปแบบปิรามิดแบบมีหลายข้อในแต่ละขั้นตอน (Multiple-item pyramidal model) ซึ่งเป็นรูปแบบที่ Krathworth, Huyser และ Linn เป็นผู้เสนอโดยการให้จำนวนข้อในแต่ละขั้นมีมากกว่าหนึ่งข้อ ทั้งนี้เพื่อให้การตอบเพื่อแยกทางได้ถูกต้องมากยิ่งขึ้น ดังภาพที่ 1



ภาพที่ 1 แสดงโครงสร้างของการทดสอบเฉพาะบุคคลรูปแบบปิรามิดแบบมีหลายข้อในแต่ละขั้น (Multiple-item Model)

จากภาพในขั้นที่ 1 ถ้าผู้สอบตอบข้อสอบถูก 0 ถึง 1 ข้อ จะแยกไปทำข้อสอบที่ 4, 5 และ 6 ถ้าตอบข้อสอบถูก 2 ถึง 3 ข้อ ก็จะแยกไปทำข้อสอบข้อที่ 7, 8 และ 9 สำหรับในขั้นตอนอื่น ๆ ก็ในทำนองเดียวกัน

กลุ่มตัวอย่างที่ใช้ในการวิจัยเป็นนักศึกษา มหาวิทยาลัยรามคำแหง ชั้นปีที่ 1 ที่ได้มาจากการสุ่มอย่างง่าย (Simple random sampling) จำนวน 2,868 คน โดยแบ่งเป็น 4 กลุ่มดังนี้

- กลุ่มที่ 1 เพื่อทดลองกลุ่มข้อสอบจำนวน 3 ฉบับ ๆ ละ 150 คน
- กลุ่มที่ 2 เพื่อทดสอบหาค่าพารามิเตอร์ของข้อสอบ 3 ฉบับ ดังนี้
 - ฉบับที่ 1 จำนวน 767 คน
 - ฉบับที่ 2 จำนวน 768 คน
 - ฉบับที่ 3 จำนวน 773 คน
- กลุ่มที่ 3 เพื่อทดลองใช้แบบสอบเฉพาะบุคคล 10 คน

กลุ่มที่ 4 เพื่อหาคุณภาพของแบบสอบถามเฉพาะบุคคล 100 คน เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัยครั้งนี้เป็นแบบสอบวิชา EN101 ของมหาวิทยาลัยรามคำแหงซึ่งผู้วิจัยเป็นผู้สร้างขึ้น โดยมีขั้นตอนในการสร้างดังนี้

1. สร้างข้อสอบตามจุดประสงค์และขอบเขตที่กำหนดขึ้น จำนวน 3 ฉบับ ๆ ละ 100 ข้อ

2. นำแบบสอบที่สร้างและปรับปรุงตามคำแนะนำของผู้เชี่ยวชาญไปทดลองใช้แล้วนำผลมาวิเคราะห์ตามทฤษฎีการวัดแบบคลาสสิก

3. คัดเลือกข้อสอบที่มีคุณภาพดีตามเกณฑ์ของทฤษฎีการวัดแบบคลาสสิก

4. นำข้อสอบที่ผ่านการวิเคราะห์แล้วมาจัดเป็นชุด ๆ ละ 90 ข้อ

5. นำแบบสอบในข้อ 4 ไปทดสอบกับกลุ่มตัวอย่าง 3 กลุ่ม ๆ ละ 767, 768 และ 773 คน ตามลำดับ เพื่อหาค่าพารามิเตอร์ของข้อสอบ แล้วคัดเลือกข้อสอบที่มีคุณภาพตามเกณฑ์ดังนี้

5.1 ค่าอำนาจจำแนกตั้งแต่ 0.50 ขึ้นไป

5.2 ค่าความยาก -3.0 ถึง +3.0

5.3 ค่าการเดาน้อยกว่า .3

อย่างไรก็ตามเพื่อให้ได้ข้อสอบที่มีค่าความยากครอบคลุมผู้สอบที่มีความสามารถสูง และผู้สอบที่มีความสามารถต่ำ ผู้วิจัยจึงได้คัดเลือกข้อสอบที่มีความยากมากกว่า 3.0 และน้อยกว่า -3.0 ไว้ด้วย โดยข้อสอบที่คัดเลือกไว้นี้มีค่าอำนาจจำแนกอยู่ระหว่าง 0.500 ถึง 0.300 ได้ข้อสอบทั้งสิ้นจำนวน 152 ข้อ

6. นำข้อสอบที่คัดเลือกตามเกณฑ์ในข้อ 5 มาจัดเรียงตามค่าความยากจากน้อยที่สุดไปหามากที่สุด แล้วจัดกลุ่มข้อสอบตามช่วงความยากตามรูปแบบปิรามิดแบบหลายข้อในแต่ละชั้น

7. ดำเนินการสร้างแบบสอบถามเฉพาะบุคคล โดยดำเนินการตามขั้นตอนดังนี้

7.1 ศึกษาทำความเข้าใจในกระบวนการทดสอบเฉพาะบุคคลโดยละเอียด ซึ่งประกอบด้วยขั้นตอนดังนี้

1. สุ่มข้อสอบจากกลุ่มข้อสอบ โดยใช้ค่าความยากมาเป็นหลักในการจัดตามรูปแบบปิรามิดแบบหลายข้อในแต่ละชั้น

2. กำหนดระดับในการเริ่มทำข้อสอบ โดยให้เริ่มจากข้อที่มีความยากปานกลาง

3. กฎในการคัดเลือกข้อสอบ จะใช้หลักว่าถ้าผู้สอบตอบข้อสอบถูกต้องข้อต่อไปจะมีค่าความยากเพิ่มขึ้น ถ้าตอบผิดข้อต่อไปก็จะมีค่าความยากลดลง

4. จัดแบบสอบถามเฉพาะบุคคล โดยข้อแรกในชั้นแรกจะใช้ข้อสอบที่มีค่าความยากเป็นค่ามัธยฐานของกลุ่ม ชั้นต่อไปก็จัดให้มีข้อสอบที่ง่ายและยากกว่าชั้นแรก เช่น ข้อ 2 จะง่ายกว่าข้อ 1 และข้อ 3 จะยากกว่าข้อ 1

5. ประมาณค่าความสามารถของผู้สอบโดยใช้หลักการของ Bayes

6. กำหนดเกณฑ์การยุติการสอบ ดังนี้

1. ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถผู้สอบน้อยกว่าหรือเท่ากับ 0.4473 หรือ

2. หากค่าความคลาดเคลื่อนสูงกว่า 0.4473 แต่ผู้ทำข้อสอบครบ 6 กลุ่ม (18 ข้อ) แล้วให้ยุติการสอบได้

7.2 นำขั้นตอนต่าง ๆ มาเขียนผังงานเพื่อให้โปรแกรมเมอร์ช่วยพัฒนาโปรแกรมให้เป็นไปตามวัตถุประสงค์ที่ต้องการ

7.3 บันทึกข้อสอบทั้งตัวคำถาม ตัวเลือกเฉลย และค่าพารามิเตอร์ของข้อสอบพร้อมคำสั่งในการควบคุมการสอบตามลักษณะโครงสร้างที่กำหนดลงในแผ่นบันทึกข้อมูล

7.4 นำโปรแกรมที่สร้างเสร็จไปทดลองใช้กับเครื่องไมโครคอมพิวเตอร์

7.5 ตรวจสอบความถูกต้องของโปรแกรมเพื่อหาข้อบกพร่องต่าง ๆ และหาทางแก้ไขปัญหาและข้อบกพร่องที่เกิดขึ้น จนกระทั่งได้โปรแกรมที่สมบูรณ์จึงนำโปรแกรมที่สร้างเสร็จไปใช้กับกลุ่มตัวอย่าง

7.6 จัดทำคู่มือการใช้โปรแกรมแบบสอบถามเฉพาะบุคคล

ผลการวิจัย

1. แบบสอบเฉพาะบุคคลแบบ ซี เอ ที วิชา EN101 ของมหาวิทยาลัยรามคำแหง มีข้อสอบจำนวน 152 ข้อ โดยแบ่งตามค่าความยากออกเป็น 11 ระดับ แต่ละระดับมีค่าความยากภายในกลุ่มต่างกันประมาณ 0.69 แบบสอบเฉพาะบุคคลชุดนี้มีค่าอำนาจจำแนกกระจายอยู่ระหว่าง 0.500 ถึง 2.681 ค่าความยากกระจายอยู่ระหว่าง -3.614 ถึง 3.997 และค่าการกระจายอยู่ระหว่าง 0.053 ถึง 0.300

2. สัมประสิทธิ์สหสัมพันธ์ระหว่างค่าความสามารถที่ประมาณได้จากแบบสอบเฉพาะบุคคลแบบ ซี เอ ที กับคะแนนจากแบบสอบแบบคลาสสิกมีค่าเท่ากับ 0.813 ซึ่งเป็นค่าที่มีนัยสำคัญทางสถิติที่ระดับ 0.013.

3. สัมประสิทธิ์สหสัมพันธ์ระหว่างค่าความสามารถที่ประมาณได้จากแบบสอบเฉพาะบุคคลแบบ ซี เอ ที กับเกรดวิชา EN101 มีค่าเท่ากับ 0.772 ซึ่งเป็นค่าที่มีนัยสำคัญทางสถิติที่ระดับ 0.1 ดังนั้นจึงอาจกล่าวได้ว่าแบบสอบเฉพาะบุคคลแบบ ซี เอ ที วิชา EN101 นี้มีความเที่ยงตรงเชิงสภาพ

สรุปและวิจารณ์ผล

1. ในการสร้างแบบสอบเฉพาะบุคคลแบบ ซี เอ ที ในครั้งนี้ผู้วิจัยได้สร้างจากกลุ่มข้อสอบที่ได้ผ่านการตรวจสอบคุณสมบัติในเรื่องการวัดเพียงมิติเดียว และความเป็นอิสระในการตอบข้อสอบโดยการวิเคราะห์องค์ประกอบ ผลการตรวจสอบพบว่าแบบสอบทั้ง 3 ฉบับ มีคุณสมบัติในการวัดเพียงมิติเดียว และมีความเป็นอิสระในการตอบข้อสอบ ซึ่งเป็นไปตามข้อตกลงเบื้องต้นของทฤษฎีของการตอบสนองต่อข้อสอบ ดังนั้นจึงสามารถนำข้อมูลจากแบบสอบทั้ง 3 ฉบับมาวิเคราะห์โดยใช้ทฤษฎีการตอบสนองต่อข้อสอบ โมเดลโลจิสติก 3 พารามิเตอร์ได้อย่างมีประสิทธิภาพ

จากการวิเคราะห์ข้อสอบโดยใช้ทฤษฎีการตอบสนองต่อข้อสอบโปรแกรม IRT(BAY) 1.0 ผู้วิจัยได้คัดเลือกข้อสอบที่มีค่าพารามิเตอร์ตามเกณฑ์ที่กำหนด ทำให้ได้กลุ่มข้อสอบ (Items pool) ที่ดีจำนวน 152 ข้อ หลังจากนั้นผู้วิจัยได้นำข้อสอบจากกลุ่มข้อสอบมาเรียง

ตามค่าความยากออกเป็น 11 ระดับ โดยมีค่าอำนาจจำแนกอยู่ระหว่าง 0.500 ถึง 2.681 ค่าความยากอยู่ระหว่าง -3.614 ถึง 3.997 และค่าการกระจายอยู่ระหว่าง 0.053 ถึง 0.300 ซึ่งจะเห็นได้ว่าข้อสอบชุดนี้มีค่าพารามิเตอร์อยู่ในเกณฑ์ที่เหมาะสมตามที่ Urry (1977) ได้เสนอแนะไว้ นอกจากนี้ค่าพารามิเตอร์ดังกล่าวยังสอดคล้องกับผลการศึกษาของ Thorndike (1982) และ Green et al. (1984)

2. ในการดำเนินการสอบโดยใช้แบบสอบเฉพาะบุคคลแบบ ซี เอ ที นั้นจะต้องดำเนินการโดยใช้เครื่องคอมพิวเตอร์ โดยผู้ใช้แบบสอบจะต้องฟังคำอธิบายในการทำข้อสอบจากผู้ดำเนินการสอบ ผู้สอบจะทำการเลือกตัวเลือกที่ต้องการ เมื่อผู้สอบตอบข้อสอบข้อแรกเสร็จ คอมพิวเตอร์จะปรากฏภาพข้อสอบข้อต่อไป ถ้าผู้สอบตอบข้อสอบชุดแรกถูกต้องตั้งแต่ 2 ข้อขึ้นไป (จาก 3 ข้อ) ข้อสอบชุดต่อไปจะมีระดับความยากเพิ่มขึ้น แต่ถ้าผู้สอบตอบข้อสอบถูกไม่เกิน 1 ข้อ (จาก 3 ข้อ) ข้อสอบชุดต่อไปจะมีระดับความยากลดลง โดยที่ผู้สอบจะไม่มีโอกาสทำข้อสอบซ้ำข้อเดิม การสอบจะดำเนินไปเรื่อย ๆ จนกว่าค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบจะมีค่าน้อยกว่าหรือเท่ากับ 0.4473 หรือผู้สอบทำข้อสอบครบ 18 ข้อ จึงจะยุติการสอบ หลังจากนั้นบนจอภาพจะปรากฏผลการสอบว่าผู้สอบทำข้อสอบทั้งหมดกี่ข้อ ถูกกี่ข้อ ใช้เวลาทำกี่นาที และความสามารถของผู้สอบเป็นเท่าไร

ข้อสังเกตที่พบในขณะที่นักศึกษากำลังสอบกับเครื่องคอมพิวเตอร์ ก็คือนักศึกษาส่วนใหญ่จะตั้งใจสอบ มีความสนใจ และกระตือรือร้นในการทำข้อสอบเป็นอย่างดี ดังนั้นจึงมีความเป็นไปได้ที่จะนำการทดสอบแบบนี้มาใช้กับนักศึกษามหาวิทยาลัยรามคำแหง ทั้งนี้เพราะมหาวิทยาลัยรามคำแหง มีศูนย์คอมพิวเตอร์ขนาดใหญ่ที่สามารถจัดสอบได้ครั้งละมาก ๆ และใช้เวลาในการสอบแต่ละครั้งเพียงเล็กน้อย (ไม่เกิน 5 นาที) ก็สามารถประมาณค่าความสามารถของผู้สอบได้ถูกต้องตรงตามสภาพความเป็นจริง นอกจากนี้การสอบโดยใช้คอมพิวเตอร์ยังมีข้อดีว่าการสอบโดยการเขียนตอบแบบเดิม (ตรึงใจ, 2534) ดังนี้

1. ใช้เวลาในการสอบน้อย แต่ให้ผลถูกต้อง
ใกล้เคียงกับความเป็นจริง

2. สามารถทราบผลการสอบทันที และสามารถ
เก็บบันทึกผลการสอบได้ด้วย

3. ผู้สอบมีทัศนคติที่ดีต่อการสอบ กล่าวคือ
ผู้สอบจะไม่เกิดความเบื่อหน่ายในการทำข้อสอบ ทั้งนี้
เพราะผู้สอบได้ทำการข้อสอบที่เหมาะสมกับระดับ
ความสามารถของตน การทำข้อสอบที่ยากเกินความ
สามารถของตนจะทำให้ผู้สอบเกิดความท้อถอยในการ
ทำข้อสอบ และมีแนวโน้มที่จะตอบข้อสอบโดยการเดา
สูง ส่วนการทำข้อสอบที่ง่ายกว่าระดับความสามารถก็
ไม่ทำให้ผู้สอบได้ใช้ความสามารถอย่างเต็มที่ ทำให้
ผู้สอบเกิดความเบื่อหน่าย ผลกระทบดังกล่าวนี้จะส่งผล
ต่อความถูกต้องแม่นยำของคะแนนลดลง และเป็นผลทำ
ให้การประมาณค่าความสามารถของผู้สอบสอบคลาด
เคลื่อนไปจากความจริง (Weiss, 1974) นอกจากนี้ยัง
สอดคล้องกับผลการวิจัยของ งามนิตย์ และปรีชา
(2527) ที่พบว่าผู้สอบส่วนใหญ่มีทัศนคติที่ดีต่อการสอบ
เฉพาะบุคคล

4. เป็นการสอบที่ยั่วผู้สอบ ทั้งนี้เพราะมี
การปรับระดับความยากของข้อสอบให้เหมาะสมกับ
ความสามารถของผู้สอบตลอดเวลา

5. นักเรียนสามารถมาสอบเมื่อไรก็ได้ที่
พร้อม โดยครูไม่จำเป็นต้องมาดำเนินการสอบด้วย
ตนเอง

3. ในการหาความเที่ยงตรงเชิงสภาพ พบว่า
คะแนนความสามารถจากแบบสอบเฉพาะบุคคลแบบ
ซี เอ ที กับเกณฑ์วิชา EN101 มีค่าสัมประสิทธิ์สหสัมพันธ์
เท่ากับ 0.772 ซึ่งสามารถสรุปได้ว่าแบบสอบเฉพาะ
บุคคลแบบ ซี เอ ที วิชา EN101 นี้มีความเที่ยงตรงเชิง
สภาพสูง เหมาะสมที่จะเป็นเครื่องมือวัดผลสัมฤทธิ์
ทางการเรียนได้อย่างมีประสิทธิภาพ นอกจากนี้ผล
การศึกษาครั้งนี้ยังสอดคล้องกับความเห็นของ Weiss
and Kingsbury (1984) ที่ว่าการสอบแบบ ซี เอ ที มี
ความเที่ยงตรงพอ ๆ กันหรืออาจมากกว่าการสอบแบบ
คลาสสิกสิ่งนี้ย่อมแสดงให้เห็นว่าการสอบแบบ ซี เอ ที
สามารถใช้แทนการสอบแบบคลาสสิกได้โดยที่ไม่ลด
ประสิทธิภาพในการประมาณความสามารถความสามารถ

ของผู้สอบแต่อย่างใด และยิ่งพบอีกว่าการสอบแบบ
ซี เอ ที ดีกว่าการสอบแบบคลาสสิกตรงที่สามารถลด
ความยาวของแบบสอบลงได้มากกว่า 50 เปอร์เซ็นต์
ข้อเสนอแนะ

1. ควรมีการวิจัยเกี่ยวกับการพัฒนาการ
ทดสอบเฉพาะบุคคลรูปแบบอื่น ๆ เช่นรูปแบบปรับ
ระดับชั้น รูปแบบแฟล็กซีเบิลเวส รูปแบบเบย์ หรือรูปแบบ
อื่น ๆ โดยทำการเปรียบเทียบประสิทธิภาพของการ
ทดสอบเฉพาะบุคคลในรูปแบบต่าง ๆ ด้วย

2. เนื่องจากทดสอบเฉพาะบุคคลแบบ ซี เอ ที
เป็นการสอบที่ให้ผู้สอบกับเครื่องคอมพิวเตอร์ซึ่ง
แตกต่างจากการสอบแบบคลาสสิกที่ให้เขียนตอบลงใน
กระดาษคำตอบ (Paper-pencil test) ดังนั้นจึงน่าจะมี
การวิจัยเพื่อประเมินผลกระทบในเชิงจิตวิทยา เช่น
ทัศนคติต่อการสอบ ความสนใจในการทำข้อสอบ
แรงจูงใจในการทำข้อสอบ และผลกระทบในด้านการ
ดำเนินการสอบ

3. เนื่องจากการประมาณค่าความสามารถของ
ผู้สอบนั้นมีหลายวิธี ดังนั้นจึงควรมีการวิจัยเพื่อเปรียบเทียบ
วิธีการในการประมาณค่าความสามารถของผู้สอบ
ว่าวิธีใดเป็นวิธีที่มีประสิทธิภาพ และให้ผลถูกต้อง
สอดคล้องกับสภาพความเป็นจริงมากที่สุด

4. ควรนำการทดสอบเฉพาะบุคคลแบบ ซี เอ ที
ที่ได้จากการวิจัยในครั้งนี้มาใช้ในการวัดผลสัมฤทธิ์
ทางการเรียนวิชา EN101 ของมหาวิทยาลัยรามคำแหง
โดยอาจเริ่มทดลองใช้ในการสอบซ่อมเพื่อเป็นโครงการ
นำร่อง (Pilot project) ก่อนก็ได้

กิตติกรรมประกาศ

งานวิจัยเรื่อง การพัฒนาการทดสอบเฉพาะ
บุคคลแบบ ซี เอ ที สำเร็จลงได้โดยได้รับความช่วยเหลือ
อย่างดียิ่งจาก รศ. ดร. คณิต ไช่มุกด์ และคุณนำชัย แซ่หลี่
ในการพัฒนาโปรแกรมสำหรับงานวิจัยครั้งนี้ ผู้วิจัยขอ
กราบขอบพระคุณเป็นอย่างสูงมา ณ โอกาสนี้ ขอขอบ
คุณมหาวิทยาลัยรามคำแหงที่ให้ทุนสนับสนุนการวิจัย
ครั้งนี้

เอกสารอ้างอิง

- งามนิตย์ ธาตุทอง และปรีชา เครือวัลย์. 2525. การให้คะแนนตัวเองในการสอบแบบทดสอบชนิดเพิลิกซีเลเวล. รายงานการวิจัย. มหาวิทยาลัยขอนแก่น.
- ตรึงใจ พูลผลอำนวย. 2534. การพัฒนาแบบสอบเฉพาะบุคคลในวิชาวิทยาศาสตร์สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3. วิทยานิพนธ์ครุศาสตรมหาบัณฑิต จุฬาลงกรณ์มหาวิทยาลัย.
- บุญชม ศรีสะอาด. 2535. การวิจัยทางการวัดผลและประเมินผล. มหาสารคาม. ภาควิชาการวัดผลและการวิจัยการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ มหาสารคาม. (อัดสำเนา).
- สงบ ลักษณะ. 2535. ความสามารถเกี่ยวกับการตอบข้อสอบ. การวัดผลการศึกษา. (พฤษภาคม-สิงหาคม) : 47-54.
- Betz, N.E. and Weiss, D.J. 1973. An Empirical Study of Computer-Administer Two-Stage Ability Testing. Research Report. Minneapolis: Psychometric Method Program, Department of Psychology, University of Minnesota.
- Green, B.F., Bock, R.D., Humphreys, L., Linn, L.G. and Reckase, M.D. 1984. Technical Guidelines for Assessing Computerized Adaptive Tests. Journal of Educational Measurement. 21 (Winter) : 347-360.
- Hambleton, R.K. and Swaminathan, H. 1985. Item Response Theory : Principles and Applications. Boston : Kluwer Nijhoff.
- Thorndike, R.L. 1982. Applied Psychometrics. Boston: Houghton Mifflin.
- Urry, V.W. 1977. Tailored Testing : A Successful Application of Latent Trait Theory. Journal of Educational Measurement. 14 (Summer): 181-196.
- Vale, C.D. and Weiss, D.J. 1975. A Simulation Study of Stradaptive Ability Testing. Research Report.
- Weiss, D.J. 1974. Strategies of Adaptive Ability Measurement. Research Report. 74(5), Department of Psychology, University of Minnesota.
- Weiss, D. J. and Kingsbury, G.G. 1984. Application of Computerized Adaptive Testing to Educational problems. Journal of Educational Measurement. 21 (Winter): 361-375.