



การจำแนกผู้เข้ารับการทดสอบโดยใช้แบบทดสอบความยาวแปรผัน The Classification Examinee by Using Variable-Length with Item Selection Testing

ปิยนุช เจริญมุล (Piyanuch Charernmool)* ปรวัดณ์ วิสูตรศักดิ์ (Porawat Visutsak)**
และ อนิราช มิ่งขวัญ (Anirach Mingkhwan)*

บทคัดย่อ

งานวิจัยนี้มุ่งเน้นด้านการพัฒนาเพื่อสร้างระบบการจำแนกผู้เข้ารับการทดสอบโดยใช้แบบทดสอบความยาวแปรผันตามกรอบแนวคิดการพัฒนาวิธีการทดสอบแบบปรับเปลี่ยนได้ด้วยคอมพิวเตอร์ซึ่งสามารถลดระยะเวลาในการทำแบบทดสอบได้จริง โดยผู้เข้ารับการทดสอบไม่ต้องทำแบบทดสอบครบตามความยาวแบบทดสอบเพื่อจัดกลุ่มผู้เข้ารับการทดสอบออกเป็นกลุ่มเก่งและกลุ่มอ่อนด้วยข้อสอบมาตรฐานปี 2551 วิชาวิทยาศาสตร์สำหรับนักเรียนช่วงชั้นที่ 3 จำนวน 140 คน พบว่าระบบที่ใช้ทดสอบมีประสิทธิภาพและมีความยืดหยุ่นมากกว่าการทดสอบแบบดั้งเดิมมีการใช้จำนวนข้อสอบเฉลี่ยเพียงร้อยละ 85.16 จากความยาวแบบทดสอบ ซึ่งพบว่าใช้จำนวนข้อสอบเฉลี่ยน้อยกว่าการทดสอบแบบดั้งเดิม 14.84 เปอร์เซ็นต์ จากการประเมินความพึงพอใจที่มีต่อระบบโดยผู้เข้ารับการทดสอบ พบว่ามีค่าเฉลี่ยความพึงพอใจอยู่ในระดับดีเท่ากับ 4.28 สามารถนำไปประยุกต์ใช้เพื่อพัฒนาระดับความรู้ของผู้เข้ารับการทดสอบกลุ่มอ่อนให้มีระดับผลการเรียนที่ดีขึ้นโดยอาจารย์ผู้สอน อีกทั้งยังเป็นการส่งเสริมให้ผู้เข้ารับการทดสอบมีเจตคติที่ดีในการสอบ

คำสำคัญ: ทฤษฎีการตอบสนองของข้อสอบ การทดสอบแบบปรับเปลี่ยนได้ด้วยระบบคอมพิวเตอร์ การคัดเลือกข้อสอบ การทดสอบแบบดั้งเดิม

Abstract

The study of the item selection for The Classification Examinee by Using Variable - Length with Item Selection

* ภาควิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีและการจัดการอุตสาหกรรม มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

** ภาควิชาวิทยาการคอมพิวเตอร์และสารสนเทศ คณะวิทยาศาสตร์ประยุกต์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

Testing aims to assess the level of knowledge of different examinees to assign the most appropriate item according to their level. The examinees doesn't test the full length test.

To reduce the duration of the testing. For until 2 groups of the examinees are classified (good and poor). Science standard test in 2008 for lower secondary students was used as a tool.

A sample group of 140 people. It was found that the testing system is more efficient and flexible than the classical testing.

The examinees are used average number of tests that 85.16 percent. When comparing the test length of system with the classical testing, a number of actual items were used fewer

14.84 percent. The Satisfactory assessment of the system by the sample group is in a good level with 4.28 of mean value.

The result can be applied not only to classify student groups for emphasizing the development of the examinees' ability who are in "poor" group to get better item result by the teacher and helping students to have better attitudes with testing.

Keyword: Item Response Theory, Item Selection, Adaptive Testing, Classical Testing.

1. บทนำ

การวัดผลของการทดสอบทางการศึกษานั้นอยู่บนพื้นฐานแนวคิดที่ว่า พฤติกรรมแสดงออกกำหนดมาจากลักษณะภายในตัวบุคคลซึ่งสังเกตไม่ได้โดยตรง ต้องอาศัยสิ่งเร้ามากระตุ้นให้บุคคลแสดงพฤติกรรมตอบสนอง คือ เครื่องมือการวัดผลสัมฤทธิ์ทางการศึกษานั้นหนึ่ง เช่น แบบทดสอบ

ใช้เป็นตัวแทนเนื้อหาการเรียนรู้ของผู้เรียน [1] หรือเรียกว่าการทดสอบแบบดั้งเดิมเพื่อวัดผลการทดสอบทางการศึกษาดังนั้นจึงกล่าวได้ว่า อีกหนึ่งเครื่องมือการวัดผลสัมฤทธิ์ทางการศึกษาคือ ระบบการจำแนกผู้เข้ารับการทดสอบโดยใช้แบบทดสอบความยาวแปรผันที่พัฒนาขึ้นตามกรอบแนวคิดวิธีการคัดเลือกข้อสอบสำหรับการทดสอบแบบปรับเปลี่ยนได้ด้วยคอมพิวเตอร์จึงเปรียบเสมือนสิ่งเร้านำไปกระตุ้นผู้เรียนให้แสดงพฤติกรรมตอบสนองเพื่อใช้ในการวัดผลการเรียนรู้ในรายวิชาวิทยาศาสตร์ การจำแนกผู้เข้ารับการทดสอบโดยใช้แบบทดสอบยาวแปรผันมีวัตถุประสงค์เพื่อประเมินความรู้ของผู้เข้ารับการทดสอบที่มีความแตกต่างกันโดยแบบทดสอบจะถูกปรับเปลี่ยนระดับความยากง่ายตามความสามารถของผู้เข้ารับการทดสอบแต่ละบุคคล ด้วยแนวความคิดที่ว่าผู้เข้ารับการทดสอบแต่ละบุคคลควรจะได้รับแบบทดสอบที่เหมาะสมกับระดับความรู้ของตนเองมากที่สุด โดยผู้เข้ารับการทดสอบที่มีความสามารถกันจะได้รับชุดข้อสอบที่แตกต่างกันและมีค่าความยากง่ายไม่เท่ากัน

การทดสอบแบบดั้งเดิมในสมัยก่อนคือ การออกแบบข้อสอบเพื่อใช้สำหรับประเมินผู้เรียนจะอยู่ในรูปแบบของกระดาษกับดินสอ (Classical Testing) ผู้เข้ารับการทดสอบจะต้องตอบคำถามทั้งหมดต่อการสอบในหนึ่งวิชา เช่น คำถามแบบเลือกตอบ 4 ตัวเลือกจำนวนทั้งสิ้น 120 ข้อ ซึ่งการทดสอบดังกล่าวมีข้อเสียที่เห็นได้อย่างเด่นชัดคือ ข้อสอบทั้ง 120 ข้อนั้น ไม่จำเป็นต้องให้ผู้สอบทำทั้งหมด เช่น ผู้สอบที่เก่ง สามารถทำข้อสอบที่ยากจำนวน 40 ข้อได้ถูกต้อง 35 ข้อ ก็น่าจะเพียงพอแล้วที่จะได้คะแนนในระดับดีมากในการสอบวิชานั้น โดยที่ไม่จำเป็นต้องเสียเวลาทำข้อสอบที่ง่ายอีก 80 ข้อที่เหลือ [1] ดังนั้นวัตถุประสงค์ของระบบการจำแนกผู้เข้ารับการทดสอบโดยใช้แบบทดสอบความยาวแปรผัน คือ การมุ่งเน้นเพื่อตอบสนองแก่การสอบที่สามารถแบ่งผู้เข้ารับการทดสอบออกเป็นกลุ่มอ่อนและกลุ่มเก่ง ซึ่งต้องการลดระยะเวลาการทำแบบทดสอบด้วยการใช้จำนวนข้อสอบให้น้อยที่สุด

งานวิจัยนี้ได้พัฒนาระบบขึ้นตามกรอบแนวคิดในการพัฒนาวิธีการทดสอบแบบปรับเปลี่ยนได้ด้วยคอมพิวเตอร์ [2] ตามทฤษฎีการตอบสนองของข้อสอบ (Item Response Theory) [3] ร่วมด้วยเทคนิคการวิเคราะห์

ข้อสอบ (Item Analysis) ระบบที่ออกแบบนี้สามารถสุ่มเลือกข้อสอบที่มีความยากง่ายในระดับต่าง ๆ ซึ่งใช้ข้อสอบมาตรฐานชั้นปี สารการการเรียนรู้พื้นฐาน วิทยาศาสตร์ช่วงชั้นที่ 3 ประจำปี 2551 เพื่อบรรจุลงในคลังข้อสอบเทียมและคลังข้อสอบจริง กำหนดให้การสิ้นสุดของการทำแบบทดสอบ [4] ให้เป็นไปตามหลักเกณฑ์การยุติตามความยาวของแบบทดสอบ (Test length) โดยใช้กฎความยาวแปรผัน (Variable-Length) [5], [6] ร่วมกับการใช้เทคนิค 25 เปอร์เซนต์ [7]

2. ทฤษฎีที่เกี่ยวข้อง

การทดสอบแบบปรับเปลี่ยนได้ด้วยระบบคอมพิวเตอร์เพื่อใช้ในการจำแนกผู้เข้ารับการทดสอบทำงานโดยการสร้างแบบทดสอบที่พิจารณาจากระดับความยากและจำนวนข้อที่สอดคล้องกับระดับความสามารถของผู้สอบ การปรับเปลี่ยนระดับความยาก หมายถึง ผู้สอบที่มีความสามารถสูงจะได้รับข้อสอบที่มีระดับความยาก ยากกว่าผู้สอบที่มีความสามารถต่ำกว่า [1] ซึ่งเป็นไปตามที่ [8] กล่าวว่า ระบบการทดสอบแบบปรับเปลี่ยนได้ ถูกพัฒนาขึ้นเพื่อให้ความเหมาะสม สำหรับการแบ่งระดับกลุ่มทักษะความสามารถของนักเรียน ในเชิงวิชาการที่มีระดับความสามารถที่แตกต่างกัน แทนการใช้งานด้วย กระดาษและดินสอในการสอบ เป็นรูปแบบใหม่ที่จะเปลี่ยนแปลงการวัดผลการศึกษาเพื่อนำไปสู่การพิจารณาโดยระบบของกลุ่มนักเรียนที่ผ่านการทดสอบสามารถ ช่วยเหลือนักเรียนได้ภายใต้เงื่อนไข โดยส่วนสำคัญที่จะช่วยให้ระบบมีความสามารถในการประเมินผลความสามารถได้เข้าใจค่าความจริงมากที่สุด คือ ส่วนของคลังข้อสอบ ที่เป็นส่วนจัดเก็บรวบรวมข้อสอบที่มีคุณภาพ โดยจะต้องเป็นข้อสอบที่ผ่านการวิเคราะห์ว่ามีคุณสมบัติของข้อสอบที่ดี นำมาประยุกต์ใช้กับกระบวนการทดสอบแบบปรับเปลี่ยนได้ ซึ่งข้อสอบแต่ละข้อจะมีระดับความยากง่ายและระดับคุณภาพของข้อสอบที่แตกต่างกันไปตามระดับพารามิเตอร์ที่กำหนด [9] ส่วนการยุติแบบทดสอบจะใช้ตามการกำหนดความยาวของแบบทดสอบซึ่งนำมาประยุกต์ใช้ตามกฎความยาวแปรผันและกำหนดระดับความยาวของแบบทดสอบตามกฎความยาวคงที่ 20, 32 และ 44 ข้อ เมื่อการทดสอบดำเนินการครบตามจำนวนข้อแล้วถือให้เป็นจุดยุติการทดสอบ [5] ร่วมกับการใช้เทคนิค 25 เปอร์เซนต์ [7] โดยกำหนดคลังข้อสอบ ให้มีขนาดตั้งแต่ 400 ข้อขึ้นไป

จะพบว่าค่าความคลาดเคลื่อนในการจัดกลุ่มผู้เข้ารับการทดสอบลดลง ตามหลักการที่ว่าผู้ที่สามารถทำแบบทดสอบ ที่ระดับยากมาก ได้ถูกต้องจะถือได้ว่าผู้นั้นสามารถทำแบบทดสอบในระดับที่ง่ายกว่าได้จึงจัดให้อยู่ในกลุ่มเก่ง และผู้ที่ไม่สามารถทำแบบทดสอบ ที่ระดับง่ายมาก ได้ถูกต้องจะถือได้ว่าผู้นั้นไม่สามารถทำแบบทดสอบในระดับที่ยากกว่าได้จึงจัดให้อยู่ในกลุ่มอ่อน [9] เนื่องจากโปรแกรมการทดสอบแบบปรับเหมาะตามระดับความสามารถของผู้สอบคือ ระบบการทดสอบด้วยคอมพิวเตอร์ที่มีการพัฒนากระบวนการสอบตั้งแต่การสร้างและการจัดเก็บข้อสอบที่เป็นคลังข้อสอบรวมถึงการทดสอบและประมวลผลโดยอาศัยคอมพิวเตอร์ [10] ในการทดสอบนั้นผู้สอบที่มีความสามารถสูงจะไม่รู้สึกเบื่อหน่ายที่จะทำข้อสอบข้อที่ง่าย ๆ จำนวนมากแบบในการสอบแบบดั้งเดิมในขณะที่ผู้สอบที่มีความสามารถต่ำกว่าก็จะไม่รู้สึกท้อแท้ที่ต้องทำข้อสอบข้อที่ยาก ๆ ผู้สอบทุกคนจะถูกวัดผลในระดับที่มีความเที่ยงตรงเท่าเทียมกันถึงแม้ว่าจะถูกวัดผลด้วยจำนวนข้อสอบที่แตกต่างกันซึ่งคุณลักษณะของระบบในข้อนี้เป็นการนำเอาเกณฑ์ มาตรฐานตามหลักจิตวิทยา มาประยุกต์ใช้ [9]

จากที่กล่าวมาทั้งหมดสามารถสรุปได้ว่า การจำแนกผู้เข้ารับการทดสอบโดยใช้แบบทดสอบความยาวแปรผันข้อสอบที่ใช้ในคลังข้อสอบจะต้องเป็นข้อสอบที่ผ่านการวิเคราะห์และได้รับมาตรฐานมีการประเมินคุณภาพและระดับความยากของข้อสอบซึ่งผู้วิจัยเลือกใช้ซึ่งใช้ข้อสอบมาตรฐานชั้นปีสาระการเรียนรู้พื้นฐานวิทยาศาสตร์ช่วงชั้นที่ 3 ประจำปี 2551 เพื่อบรรจุลงในคลังข้อสอบเทียบและคลังข้อสอบจริงซึ่งสามารถจัดระดับความยากง่ายที่แตกต่างกันไปและมีจำนวนข้อสอบที่มากเพียงพอต่อการสุ่มข้อสอบเพื่อลดโอกาสความซ้ำซ้อนที่ข้อสอบจะถูกสุ่มเลือกมาให้แก่ผู้เข้ารับการทดสอบใหม่อีกครั้งหลังจากเพิ่งได้รับข้อสอบข้อนั้นไปซึ่งเป็นส่วนหนึ่งในการสร้างแรงจูงใจในการสอบให้แก่ผู้เข้ารับการทดสอบ

3. วิธีการดำเนินงานวิจัย

3.1 การพัฒนาระบบ

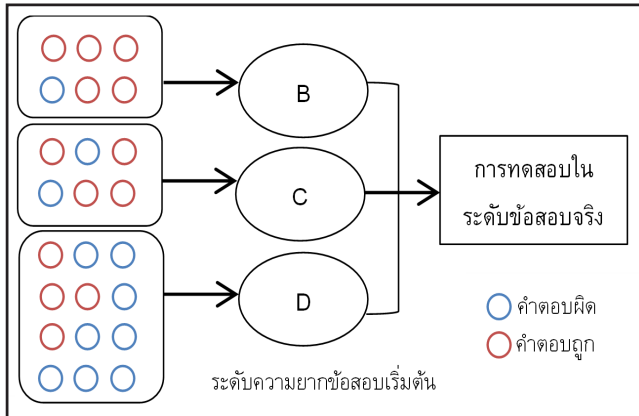
การจำแนกผู้เข้ารับการทดสอบโดยใช้แบบทดสอบความยาวแปรผันที่มีการจัดข้อสอบให้อย่างต่อเนื่องจนระดับความแม่นยำของค่าประมาณความสามารถของแต่ละบุคคล

มาถึงระดับความแม่นยำตามกำหนดการทดสอบจึงยุติกำหนดความยาวแบบทดสอบว่าจำนวนข้อสอบเท่าไรควรจัดให้แก่ผู้สอบระดับความสามารถต่างๆ เพื่อใช้ในการประเมินค่าความสามารถทางวิชาการของผู้เข้ารับการทดสอบ [11]

โดยระบบเพื่อจัดกลุ่มผู้เรียนออกเป็นกลุ่มเก่งและกลุ่มอ่อน กำหนดจุดเริ่มต้นการทดสอบโดยการประเมินผลความสามารถเริ่มต้นโดยการสุ่มใช้ข้อสอบจากคลังข้อสอบเทียบ พร้อมทั้งมุ่งเน้นวิธีการจัดสรรข้อสอบที่ใช้ในการทดสอบจริงให้มีจำนวนข้อน้อยที่สุดโดยมีเงื่อนไขว่าหากผู้เข้ารับการทดสอบสามารถทำข้อสอบในระดับ ยากมาก ถูกต้องครบ 25 เปอร์เซ็นต์ ของระดับความยาวแบบทดสอบระบบจะประเมินค่าความสามารถของผู้เข้ารับการทดสอบให้อยู่ในกลุ่มเก่ง ในทางกลับกันหากผู้เข้ารับการทดสอบทำข้อสอบในระดับง่ายมาก ผิดครบ 25 เปอร์เซ็นต์ ของระดับความยาวแบบทดสอบหรือที่เรียกว่าตามกฎความยาวคงที่ (Fixed-Length) ระบบจะประเมินค่าความสามารถของผู้เข้ารับการทดสอบให้อยู่ในกลุ่มอ่อน ซึ่งเป็นไปตามการยุติการทดสอบตามกฎความยาวแปรผัน ทำให้ผู้เข้ารับการทดสอบแต่ละคนได้รับข้อสอบจำนวนไม่เท่ากันแตกต่างกันหลายระดับความยาวแบบทดสอบจึงเกิดความคลาดเคลื่อนในการประเมินค่าความสามารถมากผู้วิจัยกำหนดความยาวแบบทดสอบเท่ากับ 32 ข้อ เพื่อให้เป็นจุดยุติการทดสอบหากผู้เข้ารับการทดสอบไม่เข้าหลักเกณฑ์การยุติการทดสอบทั้ง 2 เงื่อนไข แล้วระบบจะวิเคราะห์ค่าความสามารถด้วยวิธีการทางสถิติ เพื่อจัดว่าอยู่ในกลุ่มเก่ง หรือกลุ่มอ่อน ซึ่งการใช้เทคนิค 25 เปอร์เซ็นต์ จะช่วยลดระยะเวลาการทำแบบทดสอบ เนื่องจากผู้เข้ารับการทดสอบไม่ต้องทำแบบทดสอบจนครบระดับความยาวแบบทดสอบตามที่กำหนด

กระบวนการการทดสอบแบบปรับเปลี่ยนได้ด้วยคอมพิวเตอร์แบ่งออก 2 ส่วนคือ การทดสอบในระดับข้อสอบเทียบประกอบไปด้วยข้อสอบที่มีคุณภาพ 3 ระดับคือ พอใช้ ถึงดีมาก มีค่าระดับความยากของข้อสอบ (P) จำนวน 3 ระดับ คือ ง่าย (B) ปานกลาง (C) และยาก (D) ด้วยการสุ่มข้อสอบเทียบ จากคลังข้อสอบเทียบ จำนวน 3 ข้อ เพื่อใช้ประเมินค่าความสามารถจากการสุ่มใช้ข้อสอบเทียบด้วยวิธีการสุ่มแบบแบ่งชั้นชนิดสุ่ม (Stratified Random Sampling)

คือการแบ่งสุ่มตามระดับชั้นความยากของข้อสอบเป็นรายข้อ มีค่าความยากของข้อสอบ (P) จะมีค่าอยู่ระหว่าง 0.40 ถึง 0.60 ดังตารางที่ 1 เพื่อกำหนดจุดเริ่มต้นการทดสอบ ดังภาพที่ 1 หลังจากนั้นจะเข้าสู่การทดสอบจริงซึ่งใช้ข้อสอบจากคลังข้อสอบจริง แบ่งระดับความยากง่ายของข้อสอบออกเป็น 5 ระดับ โดยวัดจากระดับค่าความยากของข้อสอบ (P) คือ ง่ายมาก (A) ง่าย (B) ปานกลาง (C) ยาก (D) และยากมาก (E)



ภาพที่ 1 การประเมินค่าความสามารถจากการสุ่มใช้ข้อสอบเทียมเพื่อกำหนดจุดเริ่มต้นการทดสอบ

ระดับความยากของข้อสอบ หาได้โดยการหาสัดส่วนเขียนในรูปสูตร โดยแสดงรายละเอียดไว้ในตารางที่ 1

$$P = \frac{R}{N} \quad (1)$$

- เมื่อ P = ดัชนีความยาก
- R = จำนวนผู้ตอบถูกทั้งหมด
- N = จำนวนผู้เข้าสอบทั้งหมด

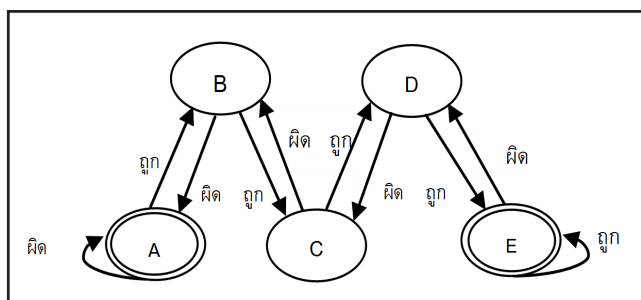
โดยการสุ่มข้อสอบจากคลังข้อสอบจริงและคลังข้อสอบเทียมจะใช้วิธีการสุ่มแบบแบ่งชั้นชนิดสุ่ม (Stratified Random Sampling) ประกอบไปด้วยข้อสอบที่มีคุณภาพ 3 ระดับ คือ พอใช้ ถึงดีมาก มีค่าความยากของข้อสอบ (P) จะมีค่าอยู่ระหว่าง 0.20 ถึง 0.80 ดังตารางที่ 1 ข้อสอบที่มีค่า P มาก ข้อสอบข้อนั้นมีคนตอบถูกมาก แสดงว่าข้อสอบง่าย ข้อสอบที่มีค่า P น้อย ข้อสอบข้อนั้นมีคนตอบถูกน้อยแสดงว่าข้อสอบยาก ข้อสอบที่มีค่า P = 0.50 เป็นข้อสอบที่มีความยากปานกลาง โดยปกติกลุ่มผู้สอบที่เก่งกว่าจะมีค่าความยากสูงกว่ากลุ่มอ่อน ค่าความยากที่เหมาะสมนั้นไม่คงที่ ทั้งนี้ขึ้นอยู่กับจุดมุ่งหมายหรือสถานการณ์ในการสอบแต่ละครั้ง [9]

ตารางที่ 1 การแปลความหมายค่าความยากของข้อสอบ

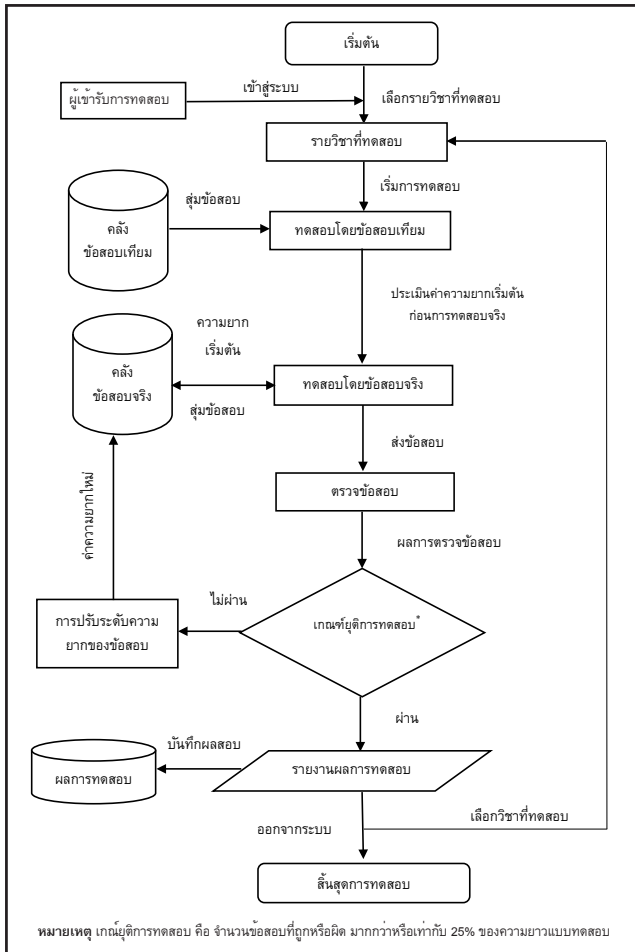
ระดับความยาก (P)	ความหมาย	คุณภาพของข้อสอบ
0.81 - 1.00	ง่ายเกินไป	ควรตัดทิ้ง
0.61 - 0.80	ง่ายมาก	พอใช้
0.51 - 0.60	ง่าย	ดี
0.50	ปานกลาง	ดีมาก
0.40 - 0.49	ยาก	ดี
0.20 - 0.39	ยากมาก	พอใช้
0.00 - 0.19	ยากเกินไป	ควรตัดทิ้ง

3.2 วิธีการทดลอง

การจัดกลุ่มผู้เข้ารับการทดสอบโดยระบบนั้น มีการวัดผลโดยการสุ่มใช้ข้อสอบเทียมดังภาพที่ 1 แล้วจะเข้าสู่การทดสอบในระดับข้อสอบจริง การปรับระดับความยากของข้อสอบ จะยึดหลักการที่ว่า ถ้าการทำข้อสอบข้อที่ผ่านมาถูกต้อง ข้อถัดไปจะยากขึ้น แต่ถ้าการทำข้อสอบข้อที่ผ่านมาผิด ข้อถัดไปจะง่ายลง [2] โดยกระบวนการนี้จะดำเนินการไปเรื่อยๆ จนกว่าจะเข้าหลักเกณฑ์การยุติการทดสอบ ซึ่งมีแผนภาพปรับระดับความยากดังภาพที่ 2 จะมีจำนวน State ทั้งหมด 5 State คือ State A (ง่ายมาก), State B (ง่าย), State C (ปานกลาง), State D (ยาก), State E (ยากมาก) กำหนดให้ State A และ State E มีสถานะเพิ่มเป็น Final State คือ สามารถยุติการทดสอบได้ทันทีเมื่อมีการทำงานภายใน Final State เป็นจำนวนมากกว่าหรือเท่ากับ 25 เปอร์เซ็นต์ของความยาวแบบทดสอบ ซึ่งเป็นจุดยุติการทดสอบตามกฎความยาวแบบแปรผัน



ภาพที่ 2 วิธีการปรับระดับความยากของข้อสอบ



ภาพที่ 3 แผนภาพการทำงานของระบบ

ตารางที่ 2 ตัวอย่างการทดสอบแบบปรับเปลี่ยนได้ของผู้รับการทดสอบ A ตามกฎความยาวแปรผัน

ข้อ	state	ผลการสอบ	state ถัดไป	คะแนนที่ได้รับ	คะแนนรวม
1	C	ผิด	B	0	0
2	B	ถูก	C	2	2
3	C	ถูก	D	3	5
4	D	ถูก	E	4	9
5	E	ผิด	D	0	9
6	D	ถูก	E	4	13
7	E	ถูก	E	5	18
8	E	ถูก	E	5	23
9	E	ถูก	-	5	28

ตารางที่ 3 ตัวอย่างการทดสอบแบบปรับเปลี่ยนได้ของผู้รับการทดสอบ B ตามกฎความยาวคงที่

ข้อ	state	ผลการสอบ	state ถัดไป	คะแนนที่ได้รับ	คะแนนรวม
1	C	ผิด	B	0	0
2	B	ถูก	C	2	2
3	C	ถูก	D	3	5
4	D	ถูก	E	4	9
5	E	ผิด	D	0	9
6	D	ถูก	E	4	13
7	E	ถูก	E	5	18
8	E	ผิด	D	0	18
9	D	ผิด	C	0	18

อธิบายตามตารางที่ 2 พบว่าผู้รับการทดสอบ A ได้รับการประมาณค่าความสามารถจากข้อสอบเทียบ ให้มีความสามารถอยู่ในกลุ่มปานกลางจะได้รับข้อสอบจริงเริ่มต้นที่ระดับปานกลาง และมีระดับการทำข้อสอบ ดังตารางที่ 2 จะเห็นได้ว่าผู้รับการทดสอบ A ระบบจะจัดให้เป็นผู้มีความสามารถอยู่ในกลุ่มเก่ง เนื่องจากการทำแบบทดสอบผู้รับการทดสอบสามารถทำข้อสอบที่อยู่ในระดับความยาก E ได้ถูกต้องครบรวม จำนวน 25 เปอร์เซ็นต์ของความยาวแบบทดสอบ (3 ข้อ) ซึ่ง State E เป็น Final State ระบบจึงจัดระดับความสามารถให้ผู้รับการทดสอบนั้นอยู่ในกลุ่มเก่งโดยไม่ต้องทำแบบทดสอบครบตามความยาวแบบทดสอบ ซึ่งใช้ไปจำนวน 9 ข้อ จากความยาวแบบทดสอบที่ระบบกำหนดความยาวแบบทดสอบเท่ากับ 10 ข้อ และจากตารางที่ 3 จะเห็นได้ว่าผู้รับการทดสอบ B เมื่อเข้ารับการทดสอบ ได้รับการประมาณค่าความสามารถจากข้อสอบเทียบ ให้มีความสามารถอยู่ในกลุ่มปานกลาง จะได้รับข้อสอบจริงเริ่มต้นที่ระดับปานกลาง และมีระดับการทำข้อสอบ ดังตารางที่ 3 พบว่ามีการทำแบบทดสอบครบตามความยาวจึงใช้การประเมินค่าความสามารถโดยเปรียบเทียบค่าตาม

สมการที่ (4) ถ้ามีค่าความจริงเป็นเท็จ ระบบจะประเมินค่าความสามารถของผู้รับการทดสอบให้อยู่ในกลุ่มอ่อน หากมีค่าความจริงเป็นจริง ระบบจะนำค่าที่ได้ไปคำนวณ

ตามสมการที่ (5)

$$Y = \text{state } (C + D + E) \quad (2)$$

$$= \text{state } (3 + 3 + 3) = 9$$

$$X = \text{state } (A + B + C) \quad (3)$$

$$= \text{state } (0 + 1 + 3) = 4$$

$$Y - X > 0 \quad (4)$$

$$9 - 4 > 0 \text{ จริง}$$

$$\text{Capability} = \frac{(Y-X) \times 100}{X} \quad (5)$$

$$= \frac{(9-4) \times 100}{4}$$

$$\text{Capability} = 125$$

เมื่อ Y คือ ผลรวมของข้อสอบที่ได้รับใน state C, D, E

X คือ ผลรวมของข้อสอบที่ได้รับใน state A, B, C

จากสมการที่ (4) มีค่าความจริงเป็นจริง จึงนำค่าที่ได้ไปคำนวณต่อในสมการที่ (5) พบว่าระดับค่า Capability เท่ากับ 125 ซึ่งมีค่ามากกว่าหรือเท่ากับ 25 ผู้เข้ารับการทดสอบ B จึงถูกจัดให้อยู่ในกลุ่มเก่ง

4. ผลการทดลอง

ผลการพัฒนาระบบการจำแนกผู้เข้ารับการทดสอบโดยใช้แบบทดสอบความยาวแปรผัน มีดังนี้



ภาพที่ 4 ภาพตัวอย่างของระบบที่พัฒนา

ผลการจัดกลุ่มผู้เข้ารับการทดสอบด้วยวิธีการคัดเลือกข้อสอบแบบปรับเปลี่ยนได้ด้วยคอมพิวเตอร์โดยใช้กฎความยาวแปรผัน จากการทดสอบโดยกลุ่มตัวอย่างจำนวน 140 คน ด้วยวิธีการสุ่มตัวอย่างแบบเจาะจง (Purposive Sampling) โดยเจาะจงคัดเลือกนักเรียนห้องที่มีระดับผลการเรียนเฉลี่ยดีที่สุดในระดับชั้นมัธยมศึกษาปีที่ 3 ได้มาจำนวน 4 ห้องเรียน จากทั้งหมด 21 ห้องเรียน จากทั้งหมด 4 โรงเรียน ที่สังกัดในเขตพื้นที่การศึกษาจังหวัดกาญจนบุรี

มีผลการจัดกลุ่มผู้เข้ารับการทดสอบ ดังนี้

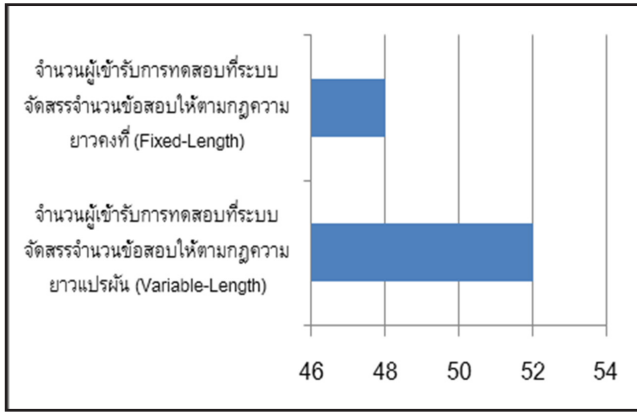
ตารางที่ 4 ผลการจัดกลุ่มผู้เข้ารับการทดสอบ

กลุ่มตัวอย่าง	จำนวนข้อสอบที่ใช้จริง		เก่ง	อ่อน
	ค่าเฉลี่ย	ส่วนเบี่ยงเบนมาตรฐาน		
โรงเรียนที่ 1	27.70	6.45	3	31
โรงเรียนที่ 2	26.67	6.47	4	36
โรงเรียนที่ 3	24.58	6.81	2	28
โรงเรียนที่ 4	30.03	4.92	1	35
รวม	27.25	6.16	10	130
เปอร์เซ็นต์	85.16		7	93

จากตารางที่ 4 พบว่าระบบที่ใช้ทดสอบมีประสิทธิภาพและมีความยืดหยุ่นมากกว่าการทดสอบแบบดั้งเดิมจริง โดยกลุ่มตัวอย่างผู้เข้ารับการทดสอบมีการใช้จำนวนข้อสอบเฉลี่ยเพียง 85.16 เปอร์เซ็นต์ และมีส่วนเบี่ยงเบนมาตรฐาน 6.16 ซึ่งใช้จำนวนข้อสอบเฉลี่ยน้อยกว่าการทดสอบแบบดั้งเดิมที่ต้องทำข้อสอบครบ 100 เปอร์เซ็นต์ ลดลงถึง 14.84 เปอร์เซ็นต์ โดยการเปรียบเทียบ จำนวนผู้เข้ารับการทดสอบที่ระบบจัดสรรจำนวนข้อสอบให้ตามกฎความยาวแปรผัน (Variable - Length) มีจำนวนถึง 52 เปอร์เซ็นต์ ซึ่งมากกว่าจำนวนผู้เข้ารับการทดสอบที่ระบบจัดสรรจำนวนข้อสอบให้ตามกฎความยาวคงที่ (Fixed - Length) ที่มีเพียง 48 เปอร์เซ็นต์ ดังกราฟแสดงในภาพที่ 5 ซึ่งเพียงพอที่จะใช้ประมาณค่าความสามารถของผู้เข้ารับการทดสอบเพื่อจัดกลุ่มผู้เข้ารับการทดสอบออกเป็นกลุ่มอ่อนและกลุ่มเก่ง โดยมีค่าเฉลี่ยความพึงพอใจ เท่ากับ 4.28 และส่วนเบี่ยงเบนมาตรฐาน เท่ากับ 0.72

5. สรุปผล

จากผลการทดสอบสามารถนำไปประยุกต์ใช้ในการจัดกลุ่มผู้เรียนเพื่อเน้นสำหรับพัฒนาระดับความรู้ความสามารถของผู้เข้ารับการทดสอบที่อยู่ในกลุ่มอ่อนให้มีระดับผลการเรียนที่ดีขึ้น โดยอาจารย์ผู้สอนประจำวิชา อีกทั้งยังเป็น การส่งเสริมให้เกิดการพัฒนาวิธีการเรียนการสอนให้มีประสิทธิภาพเหมาะสมสูงสุดแก่ผู้เรียนที่มีความสามารถแตกต่างกัน การทดสอบด้วยระบบดังกล่าวยังช่วยให้ผู้เข้ารับการทดสอบไม่รู้สึกเบื่อหน่ายต่อการสอบ



ภาพที่ 5 กราฟแสดงผลการประมาณค่าความสามารถของผู้เข้ารับการทดสอบได้โดยใช้จำนวนข้อสอบน้อยกว่า

6. เอกสารอ้างอิง

- [1] C. Kamolsin and P. Visutsak, *A practical computerized adaptive testing*. Department of Information Technology, Faculty of Industrial Technology and Management, KMU, pp. 8, 2012.
- [2] P. Visutsak. "A Practical Items Selection for Web-Based Adaptive Testing." *In Proceedings of the 1st International Conference on Technical Education (ICTechEd2013)*, Faculty of Technical Education (ICTechEd2013), pp. 146-149, 2013.
- [3] P. Charernmool and P. Visutsak, "A Framework for the Development of Computerized Adaptive Testing." *The Fifth TCU International e-learning Conference 'Overcome the Uncertainty of Technology in Education*, Hotel Windsor Suites & Convention on Sukhumvit 20 Bangkok, 6-7 Aug 2014.
- [4] Y-C. Ho Yen, W-W. R. G. Liao and L-J Chen. "Reducing the Impact of Inappropriate Items on Reviewable Computerized Adaptive Testing." *Educational Technology & Society*, Vol. 15, No. 2, pp. 231-243, 2012.
- [5] สุดารัตน์ หวลมุกดา. *ประสิทธิภาพของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจ*
- [6] จำแนกภายใต้เงื่อนไขต่างกันโดยใช้วิธีการจำลองข้อมูล. *ปริญญาโทพนธ์การศึกษาดุษฎีบัณฑิต สาขาวิชาการทดสอบและวัดผลการศึกษา บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ*, 2550.
- [7] ศิริชัย กาญจนวสี. *ทฤษฎีการทดสอบแนวใหม่*. กรุงเทพฯ. จุฬาลงกรณ์มหาวิทยาลัย, 2555.
- [8] รังสรรค์ ไกรสรานนท์. *การวิเคราะห์ข้อสอบด้วยเทคนิค 25 %*. สืบค้นเมื่อวันที่ 10 พฤษภาคม 2558 จาก http://www.tni.ac.th/web/upload/files/article/article_25.pdf.
- [9] J. Park, S. Lee and P. Kim. "Web-based adaptive testing system (wats) for classifying students academic bility." *Turkish Online Journal of Distance Education-TOJDE*. October 2012 : 13. pp. 4 Article-2, 2012.
- [10] H-P. Wang, B-C. Kuo, Y-H. Tsai, and C-H. Liao. "A cefr-based computerized adaptive testing system for chinese proficiency." *The Turkish Online Journal of Educational Technology*. Vol. 11 Issue 4. Turkish, 2012.
- [11] สิริลักษณ์ เกษรปทุมานันท์. *การเปรียบเทียบความตรงตามสภาพในการประมาณค่าความสามารถของผู้สอบจากการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ ที่ใช้เกณฑ์การคัดเลือกข้อสอบ ชั้นแรก อัตราการใช้ข้อสอบซ้ำ และเกณฑ์ยุติการทดสอบที่ต่างกัน*. วิทยานิพนธ์ครุศาสตรมหาบัณฑิต สาขาวิชาการวัดและประเมินผลการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย, 2549.
- [12] สุนันท์ พลอาษา. *การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบที่ดำเนินการทดสอบบนเว็บเพจ กลุ่มสาระการเรียนรู้ภาษาไทย เรื่อง ความรู้พื้นฐานทางภาษาและการพัฒนาทักษะทางภาษา ชั้นมัธยมศึกษาปีที่ 3*. วิทยานิพนธ์การศึกษามหาบัณฑิต สาขาวิชาการวัดผลการศึกษา มหาวิทยาลัยมหาสารคาม, 2551.