

ระบบการวิเคราะห์ข้อสอบออนไลน์ด้วยทฤษฎีการตอบสนองข้อสอบ แบบอัตโนมัติ*

Automatic Online Item Analysis System by Item Response Theory

สุโกศล วโนทยาพิทักษ์**

บทคัดย่อ

ระบบการวิเคราะห์ข้อสอบออนไลน์ด้วยทฤษฎีการตอบสนองข้อสอบแบบอัตโนมัติเป็นการนำข้อมูลที่ได้จากการสอบออนไลน์มาเพิ่มคุณค่าโดยนำมาวิเคราะห์คุณภาพข้อสอบตามทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) ซึ่งเป็นทฤษฎีที่ว่าด้วยความสัมพันธ์ระหว่างความสามารถแฝงภายในบุคคล (Trait) กับพฤติกรรมกรรมการตอบสนองข้อสอบของบุคคลในรูปความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องระบบนี้พัฒนาในรูปแบบเว็บแอปพลิเคชันโดยใช้หลักการวิเคราะห์และออกแบบระบบเชิงวัตถุ และเครื่องมือในการพัฒนาแบบเปิดเผยรหัส (Open source) ในส่วนของการวิเคราะห์ข้อสอบด้วยทฤษฎีการตอบสนองข้อสอบจะเริ่มต้นแบบอัตโนมัติทันทีที่การสอบเสร็จสิ้นลงโดยระบบจะนำเอาคะแนนผลการสอบจากฐานข้อมูลมาวิเคราะห์คุณภาพข้อสอบ 3 พารามิเตอร์ด้วยวิธีการมาร์จินัลแมกซ์ลิคฮูด (Marginal Maximum Likelihood Method) ซึ่งเป็นอัลกอริทึมการคาดหวัง-การหาจุดสูงสุด (Expectation-Maximization: E-M) ได้แก่ ค่าอำนาจจำแนก ค่าความยากง่าย

ค่าความง่ายต่อการเดาถูกของผู้สอบที่มีความสามารถต่ำ และจะนำมาคำนวณรวมกันเพื่อหาค่าฟังก์ชันสารสนเทศ ซึ่งเป็นการวัดความเชื่อมั่นข้อสอบทั้งฉบับข้อเสนอแนะในการนำการวิจัยไปใช้ คือ (1) จำนวนข้อสอบควรมีจำนวนมากกว่าการสอบปกติด้วยกระดาษในเวลาสอบที่เท่ากันเพราะการสอบออนไลน์ผู้สอบตอบคำตอบได้เร็วกว่า (2) ในแต่ละข้อมีความยากง่ายต่อการเดาข้อสอบไม่เท่ากันจึงอาจปรับคะแนนตามน้ำหนักความยากง่ายต่อการเดาข้อสอบ เพื่อจะได้คะแนนสอบที่แท้จริงของผู้สอบ (3) ยังสามารถประยุกต์การวิจัยด้วยแนวทางอื่นเช่น ข้อสอบแบบปรับตัวได้ด้วยคอมพิวเตอร์ (CAT) ทฤษฎีการตอบสนองข้อสอบหลายมิติ (mIRT) (4) ในการเปรียบเทียบทฤษฎีการตอบสนองข้อสอบการทฤษฎีการสอบแบบดั้งเดิมควรใช้กลุ่มตัวอย่างตั้งแต่ 200-500 คนในกรณีทฤษฎีการตอบสนองข้อสอบ และ 500-1,000 คน ในกรณีทฤษฎีดั้งเดิม (5) อัลกอริทึม E-M ที่ใช้ในการวิจัยนี้ใช้เวลาประมวลผลนานมาก ดังนั้นงานวิจัยในครั้งต่อไปควรหาอัลกอริทึม หรือแนวทางการประมวลผลอื่นๆ มาใช้เพื่อลดเวลาการประมวลผล

*ได้รับทุนสนับสนุนการวิจัยจากวิทยาลัยเซาธ์อีสท์บางกอกปีการศึกษา 2558

**ผู้ช่วยศาสตราจารย์ประจำคณะวิทยาศาสตร์และเทคโนโลยี วิทยาลัยเซาธ์อีสท์บางกอก

Abstract

The automatic online item analysis system by Item Response Theory (IRT) is taking the online testing data to analyze item automatically with IRT, the theory about the relationship of the personal ability (trait) and the item responsive behavior in the form of correct answer probability. This system is developed in the web-based application form by uses the principle of object-oriented analysis and design and developing open source tools. In the part of item analysis with IRT will automatically start up when a test is finished by taking item scores from the database to process with marginal maximum likelihood method by expectation-maximization algorithm for three parameters item quality, discrimination, a difficulty, and a guessing right by low ability examinee and the last process is integrated three parameters to calculate an information

function. The suggestion for implementing this research is (1) the number of items could be much more than the regular paper testing, because of the tester can response item faster (2) in each item has difference guessing factor, so there might be adjusted score according to guessing factor for a real score of tester (3) there have another approach of research to apply such as Computerized Adaptive Testing (CAT), multidimensional Item Response Theory (mIRT) (4) for comparing the classical theory to the item response theory could be use sample of tester between 200-500 in the case of item response theory and 500 - 1,000 in the case of classical theory and (5) the E-M algorithm in this research take a long time of processing, so the next research could be use another algorithm or approach for reducing the processing time.

บทนำ

การนำระบบการสอบออนไลน์ผ่านเครือข่ายเข้ามาใช้แทนที่การสอบแบบดั้งเดิมที่ใช้กระดาษกำลังได้รับความนิยมแพร่หลายเพราะสามารถลดเวลาและค่าใช้จ่ายตลอดจนอำนวยความสะดวกในการบริหารจัดการข้อสอบให้แก่ผู้ออกข้อสอบและผู้สอบ อย่างไรก็ตามระบบการสอบออนไลน์ส่วนใหญ่จะสิ้นสุดกระบวนการเมื่อนักศึกษาเข้าสอบและระบบรวมคะแนนสอบของนักศึกษาเพียงเท่านั้น และคาดหวังเพียงแค่คะแนนสอบ แต่มองข้ามข้อมูลอันเป็นประโยชน์ที่สามารถนำไปต่อยอดเพิ่มคุณค่าในเรื่องการวิเคราะห์หาคุณภาพของข้อสอบเพื่อนำมาปรับปรุงและใช้งานใหม่ได้ในอนาคต

แม้ว่าบางครั้งอาจเกิดความตระหนักในการนำข้อมูลที่ได้จากการสอบออนไลน์มาวิเคราะห์หาคุณภาพของข้อสอบ แต่ระบบส่วนใหญ่ก็ไม่ได้เอื้ออำนวยความสะดวกเพราะอยู่นอกขอบเขตของระบบทำให้เกิดภาระหนักต่อผู้วิเคราะห์

ข้อสอบที่ต้องแปลงฐานข้อมูลให้อยู่ในรูปแบบโปรแกรมกระดานทำการ (Spreadsheet) แล้วนำมาใส่สูตรการคำนวณด้วยมือ (Manual) ซึ่งมีโอกาสผิดพลาดสูง และเมื่อได้ผลการวิเคราะห์มาแล้วจึงค่อยนำไปปรับปรุงข้อสอบในระบบอีกครั้งหนึ่ง ด้วยความซับซ้อนและขั้นตอนมากมายทำให้เกิดความย่อท้อที่จะนำผลสอบมาวิเคราะห์ ซึ่งผลสุดท้ายผู้ใช้ระบบการสอบออนไลน์ส่วนใหญ่ก็มักจะละเลยการวิเคราะห์ข้อสอบต่างๆ ที่เป็นกระบวนการที่สำคัญของหลักการวัดผลการศึกษาที่ดี

ระบบการวิเคราะห์ข้อสอบออนไลน์ด้วยทฤษฎีการตอบสนองข้อสอบแบบอัตโนมัตินี้จากการวิจัยนี้จะช่วยแก้ปัญหาดังกล่าวโดยนำข้อมูลการสอบมาใช้ในการวิเคราะห์หาคุณภาพข้อสอบตามหลักการวัดผลโดยอาศัยทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) (ศิริชัย กาญจนวาสี, 2555) ที่มีข้อดีกว่าทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory: CTT) และระบบนี้

ยังเื้อออำนวยความสะดวกต่อผู้ออกข้อสอบให้สามารถคัดแยกข้อสอบตามผลที่ได้จากการวิเคราะห์เพื่อสร้างคลังข้อสอบ (Item Bank) ที่สามารถนำมาใช้ได้อีกในอนาคต

จากการศึกษางานวิจัยเกี่ยวกับระบบการวิเคราะห์ข้อสอบด้วยคอมพิวเตอร์พบว่า มีผลงานวิจัยทั้งภายในประเทศและต่างประเทศ อาทิ การพัฒนาระบบการวิเคราะห์และการสร้างคลังข้อสอบผ่านเครือข่ายอินเทอร์เน็ตสำหรับมหาวิทยาลัยเอกชน (กิตติมา เจริญศิริ, 2550) การวิเคราะห์คุณภาพข้อสอบด้วยโปรแกรมคอมพิวเตอร์ (จักรศิริ ปิยะพิมลสิทธิ์, 2548). The Intelligent Assessment System in Web-based Distance Learning Education (RuiminShen, Yiyang Tang, Tong Zhen Zhang, 2001) และ A Novel Web-based Online Examination System for Computer Science Education (Yuan Zhenming, Zhang Liang, Zhan Guohua, 2003)

การวิจัยในครั้งนี้เป็นการต่อยอดจากงานวิจัยก่อนหน้านี้ของผู้วิจัยเอง คือ การพัฒนาระบบการสอบออนไลน์ผ่านเครือข่ายอินเทอร์เน็ต (สุโกศล วโนทยาพิทักษ์, 2551) ซึ่งเป็นการพัฒนาระบบการสอบออนไลน์ธรรมดาที่ยังไม่ได้มีการวิเคราะห์ข้อสอบ การพัฒนาระบบคลังข้อสอบออนไลน์แบบมีวงจรการวิเคราะห์ข้อสอบอัตโนมัติ (สุโกศล วโนทยาพิทักษ์, 2555) ซึ่งเป็นการเพิ่มส่วนของการวิเคราะห์ข้อสอบอัตโนมัติโดยใช้ทฤษฎีการทดสอบแบบดั้งเดิม และการวิจัยครั้งนี้ได้นำเอาทฤษฎีการสอบแบบตอบสนองข้อสอบมาวิเคราะห์ข้อสอบแทนทฤษฎีการทดสอบแบบดั้งเดิม

วัตถุประสงค์และขอบเขตการวิจัย

1. เครื่องมือที่ใช้ในการพัฒนาระบบทั้งหมดของงานวิจัยนี้ผู้วิจัยเลือกใช้ซอฟต์แวร์เปิดเผยแพร่ (Open source) เท่านั้น เพื่อหลีกเลี่ยงข้อจำกัดด้านลิขสิทธิ์ และเป็นแนวทางให้ผู้ที่จะวิจัยหัวข้อนี้ในอนาคต

2. กลุ่มตัวอย่างในการวิจัยเป็นนักศึกษาระดับปริญญาตรีวิทยาลัยเซาธ์อีสท์บางกอกจำนวน 25 คน และจำนวนข้อสอบที่ใช้ในการวิจัย 7 รายวิชา

หลักการวิเคราะห์ข้อสอบ

การวิเคราะห์ข้อสอบมี 2 ลักษณะ คือ วิเคราะห์เป็นรายข้อและวิเคราะห์ทั้งฉบับ วิธีการวิเคราะห์สามารถทำได้หลายแนวทางขึ้นกับทฤษฎีการวัดผลที่นำมาใช้โดยทั่วไปมีอยู่ 2 แนวทางหลัก ได้แก่ ทฤษฎีการทดสอบแบบดั้งเดิม (Classical Testing Theory: CTT) และทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) ทฤษฎีการทดสอบแบบดั้งเดิมอยู่บนพื้นฐานของตัวแบบการวัดและข้อตกลงเบื้องต้นเกี่ยวกับแบบทดสอบคู่ขนาน (Allen and Yen, 1979 และ Linn, 1989) โดยคะแนนที่ได้จากการวัดของผู้สอบแต่ละคนเป็นผลรวมของคะแนนจริง (True score) และคะแนนความคลาดเคลื่อน (Error score) ของผู้สอบนั้นแต่เป้าหมายของการวัด คือ ต้องการให้คะแนนที่วัดได้มีค่าใกล้เคียงความสามารถที่แท้จริงของผู้สอบมากที่สุด จึงต้องหาวิธีการเพื่อให้เกิดความคลาดเคลื่อนน้อยที่สุดเท่าที่จะทำได้

ตามทฤษฎีนี้การวิเคราะห์ข้อสอบเป็นรายข้อจะวิเคราะห์ค่าความยาก (Difficulty) ค่าอำนาจจำแนก (Discrimination) และประสิทธิภาพของตัวลวง (Effectiveness of distracters) ส่วนการวิเคราะห์ข้อสอบทั้งฉบับ จะวิเคราะห์ค่าความตรง หรือความเที่ยงตรง (Validity) ค่าความเที่ยงหรือความเชื่อมั่น (Reliability) ของแบบทดสอบ

อย่างไรก็ตามทฤษฎีการทดสอบแบบดั้งเดิมยังมีข้อจำกัดเนื่องจากมีพื้นฐานของการวัดผลแบบอิงกลุ่ม ทำให้ค่าสถิติแปรเปลี่ยนไปตามกลุ่มผู้สอบแต่การวัดความสามารถของผู้สอบกลับขึ้นกับอยู่กับค่าสถิติเหล่านั้นทำให้ผลการวิเคราะห์ไม่แน่นอน ประการที่สอง คือ การวิเคราะห์แบบนี้กำหนดให้ค่าความคลาดเคลื่อนมาตรฐานของผู้สอบทุกคนเท่ากันซึ่งโดยธรรมชาติแล้วเป็นสิ่งที่เป็นไปได้ ประการสุดท้าย คือ การสร้างข้อสอบคู่ขนานด้วยทฤษฎีนี้ทำได้ยากเพราะพื้นฐานจากการวัดผลแบบอิงกลุ่มที่ค่าสถิติของแต่ละกลุ่มแตกต่างกัน

จากข้อจำกัดของทฤษฎีแบบดั้งเดิมทำให้เกิดแนวทางใหม่ที่เรียกว่าทฤษฎีการตอบสนองข้อสอบ โดยหลักการของทฤษฎีนี้การวิเคราะห์ข้อสอบรายข้อจะวิเคราะห์ค่าความยาก ค่าอำนาจจำแนก ค่าความน่าจะเป็นของการเดาถูกจากผู้สอบที่มีความสามารถต่ำ ค่าฟังก์ชันสารสนเทศ

ของข้อสอบ (Item Information Function) ส่วนการวิเคราะห์ข้อสอบทั้งฉบับจะวิเคราะห์ค่าฟังก์ชันสารสนเทศของแบบทดสอบ (Test Information Function) และความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถซึ่งถ้าความคลาดเคลื่อนมาตรฐานของการประมาณ

ค่ามีค่าต่ำ แสดงว่าการประมาณค่าความสามารถของผู้เรียนมีความแม่นยำสูงในการวิจัยครั้งนี้ ผู้วิจัยเลือกใช้โมเดลทฤษฎีการสอบแบบตอบสนองข้อสอบแบบ 3 พารามิเตอร์ ดังสมการต่อไปนี้

$$P_i(\theta, a_i, b_i, c_i) = c_i + (1 - c_i) \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}} \dots (1)$$

และฟังก์ชันสารสนเทศดังต่อไปนี้

$$I_j(\theta_j) = \sum_i I_{ij}(\theta_j, b_i, a_i, c_i) = \sum_i a_i^2 \frac{Q(\theta) \left[\frac{P(\theta) - c}{1 - c} \right]^2}{P(\theta)} \dots (2)$$

โดยที่ $i = 1, 2, 3, \dots, n$

θ = ระดับความสามารถที่แท้จริง

$P_i(\theta)$ = ความน่าจะเป็นของผู้สอบที่มีความสามารถในระดับ θ จะตอบข้อสอบข้อที่ i ได้ถูกต้อง

$Q_i(\theta)$ = ความน่าจะเป็นของผู้สอบที่มีความสามารถในระดับ θ จะตอบข้อสอบข้อที่ i ไม่ถูกต้อง

$I_j(\theta_j)$ = ฟังก์ชันสารสนเทศ

a_i = ค่าอำนาจจำแนกของข้อสอบข้อที่ i

b_i = ค่าความยากของข้อสอบข้อที่ i

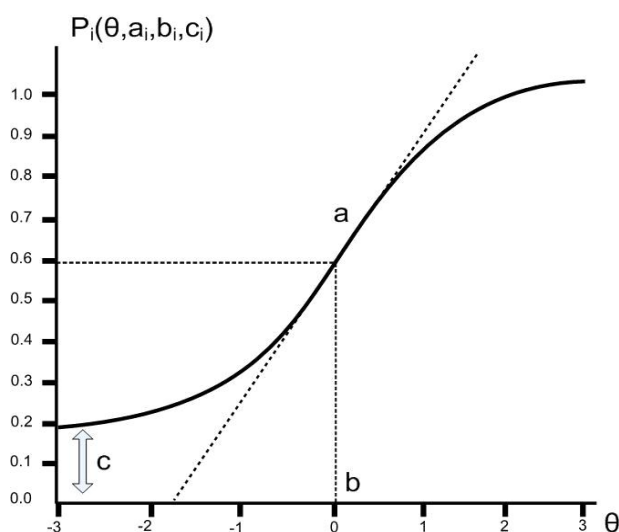
c_i = ค่าความง่ายต่อการเดาถูกโดยผู้สอบที่มีความสามารถต่ำ

D = ค่าปรับสเกลซึ่งมีค่าประมาณ 1.7

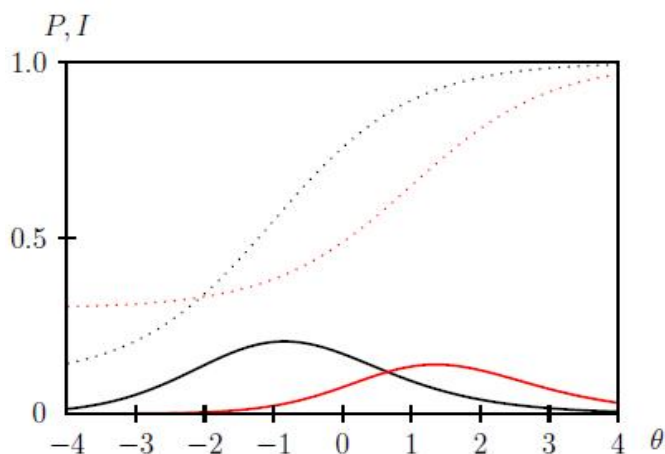
e = ค่าคงที่ซึ่งมีค่าประมาณ 2.71828

ในการประมาณค่าพารามิเตอร์ทั้งสาม ผู้วิจัยเลือกใช้วิธีการมาร์จินัลแมกซ์ิมัมไลเคิลฮูด (Marginal Maximum Likelihood) ของ Bock และ Aitkin (1981) ซึ่งจัดเป็น

อัลกอริทึมการคาดหวัง-การหาจุดสูงสุด (Expectation-Maximization; E-M)



รูปที่ 1 ฟังก์ชันการตอบสนองข้อสอบแบบ 3 พารามิเตอร์



รูปที่ 2 นำฟังก์ชันการตอบสนองข้อสอบและฟังก์ชันสารสนเทศมาพล็อตให้อยู่ในเฟรมเดียวกัน

ในรูปกราฟเส้นประ 2 เส้นที่อยู่ด้านบนแสดงถึงฟังก์ชันการตอบสนองของข้อสอบ 2 ข้อ ซึ่งเมื่อนำพารามิเตอร์จากรูปที่ 1 มาคำนวณแล้วจะทำให้ได้ฟังก์ชันสารสนเทศเป็นกราฟเส้นทึบ 2 เส้น ด้านล่างที่สอดคล้องกับข้อสอบแต่ละข้อโปรแกรมจะทำการคำนวณค่า

พารามิเตอร์ในสมการ (1) ก่อน หลังจากนั้นจึงนำค่าที่ได้จากสมการ (1) มาทำการคำนวณหาฟังก์ชันสารสนเทศในสมการ (2) ในส่วนของขั้นตอนการคำนวณจะแสดงด้วยตารางดังต่อไปนี้

		Parameters		
		a	b	c
Items	1	1.25	-0.5	0
	2	1.5	0	0.25
	3	1.25	1	0
	4	1	1.5	0

รูปที่ 3 ตัวอย่างผลการคำนวณ 3 พารามิเตอร์

		θ						
		-3	-2	-1	0	1	2	3
Items	1	4.91E-03	3.96E-02	2.57E-01	7.43E-01	9.60E-01	9.95E-01	9.99E-01
	2	4.76E-04	6.06E-03	7.24E-02	5.00E-01	9.28E-01	9.94E-01	1.00E+00
	3	2.03E-04	1.70E-03	1.41E-02	1.07E-01	5.00E-01	8.93E-01	9.86E-01
	4	4.76E-04	2.60E-03	1.41E-02	7.24E-02	2.99E-01	7.01E-01	9.28E-01

รูปที่ 4 ค่า θ ที่ได้จากการคำนวณในแต่ละข้อ

θ	-3	-2	-1	0	1	2	3
L	9.63E-08	4.22E-06	1.36E-04	9.92E-04	4.30E-04	1.86E-05	2.56E-07

รูปที่ 5 การค้นหาค่า θ ที่มากที่สุดกับ likelihood function

		I(θ)						
		-3	-2	-1	0	1	2	3
Items	1	0.00054	0.00451	0.03724	0.27681	1.05286	0.63374	0.10444
	2	0.00000	0.00000	0.00006	0.00818	0.43939	0.77814	0.10136
	3	0.00054	0.00451	0.03724	0.27681	1.05286	0.63374	0.10444
	4	0.00321	0.01741	0.09032	0.37745	0.72250	0.37745	0.09032

รูปที่ 6 ค่าฟังก์ชันสารสนเทศข้อสอบตามระดับ θ

แนวทางการพัฒนาระบบ

โดยใช้หลักการวิเคราะห์และออกแบบเชิงวัตถุ (Object-oriented Analysis and Design) (AtulKahate, 2004) มาพัฒนาเว็บแอปพลิเคชัน (Williams H.E., 2004)

ในส่วนเครื่องมือพัฒนาระบบผู้วิจัยเลือกซอฟต์แวร์เปิดเผยแพร่ (Open source Software) ได้แก่ ระบบฐานข้อมูล MySQL ภาษาสคริปต์ฝั่งเซิร์ฟเวอร์ PHP ภาษาสคริปต์ฝั่งไคลเอนท์ JavaScript เว็บเซิร์ฟเวอร์ Apache และโปรแกรมช่วยเขียนโค้ด NetBeans IDE

วิธีการพัฒนาระบบ

การวิจัยครั้งนี้ใช้หลักการวิเคราะห์และออกแบบเชิงวัตถุ (Object-oriented Analysis and Design) โดยใช้แผนภาพยูเอ็มแอล (Unified Modeling Language; UML) และแผนภาพความสัมพันธ์ระหว่างเอนทิตี (Entity Relationship Diagram; ERD) เพื่อออกแบบฐานข้อมูล โดยมีขั้นตอนการพัฒนาดังนี้ คือ

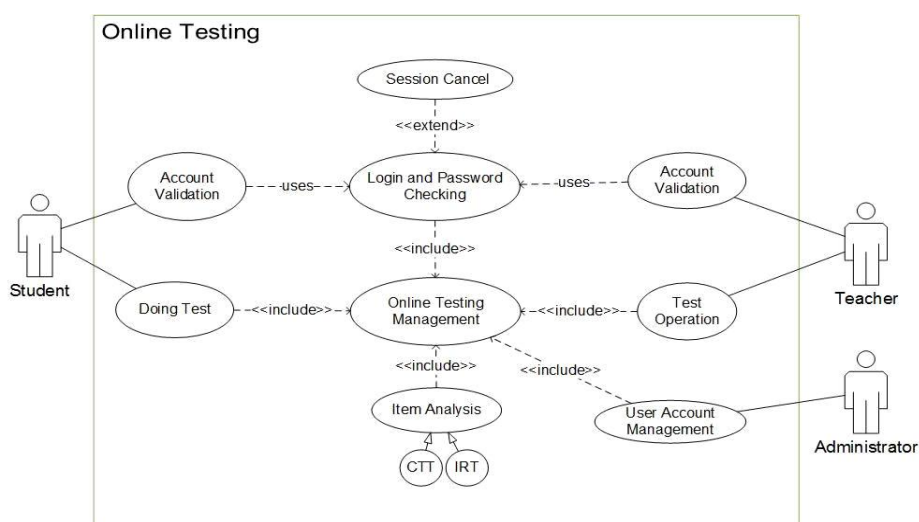
1. ขั้นตอนการวิเคราะห์ระบบ

1.1 แผนภาพสถิตย (Static Object-oriented Diagram) ประกอบด้วยแผนภาพส่วนประกอบและความสัมพันธ์ (Use Case Diagram) แผนภาพคลาส (Class Diagram) และแผนภาพดีพลอยเมนต์ (Deployment Diagram)

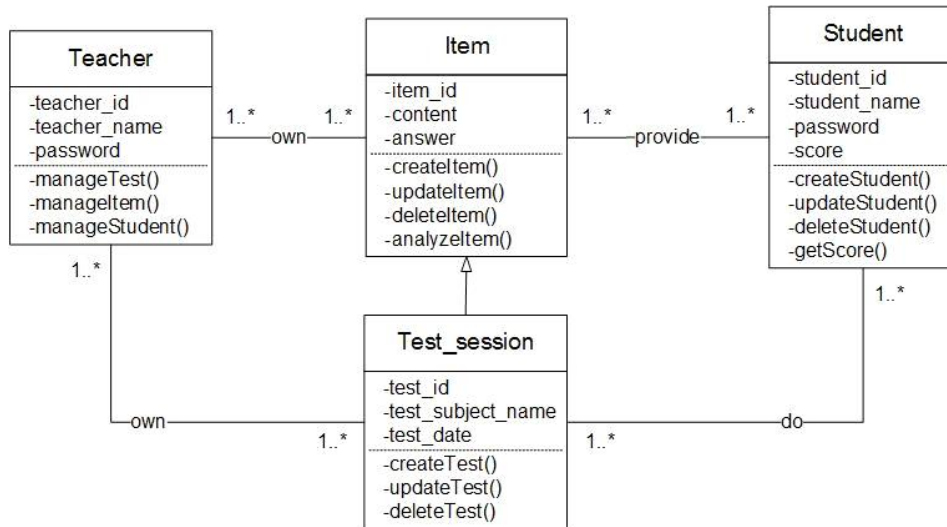
1.2 แผนภาพพลวัต (Dynamic Object-oriented Diagram) ประกอบด้วย แผนภาพลำดับ (Sequence Diagram) และแผนภาพสถานะ (State Diagram)

2. ขั้นตอนการออกแบบระบบ

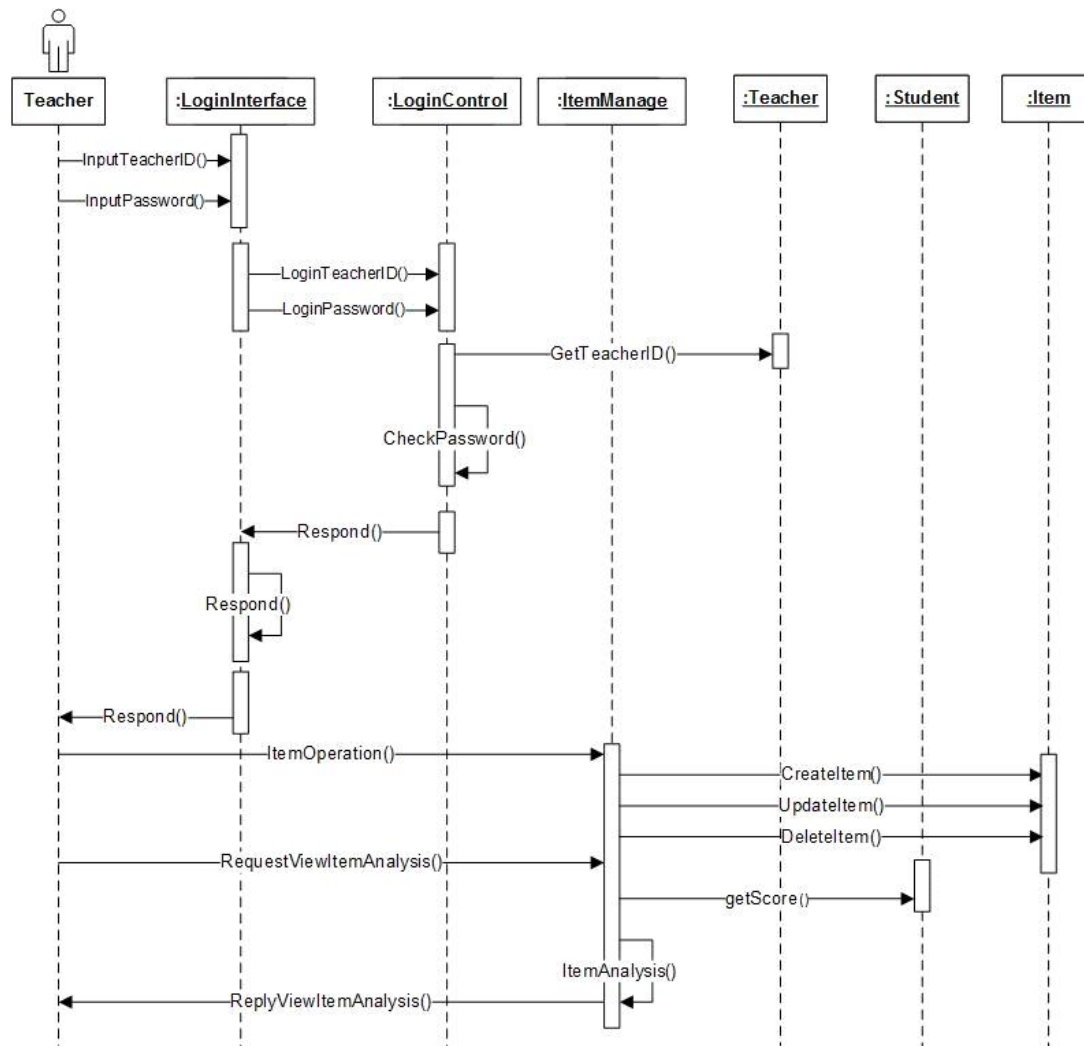
1. ปรับปรุงแผนภาพ (Refinement)
2. ออกแบบสถาปัตยกรรมของแอปพลิเคชัน (Application Architecture Design)
3. ออกแบบสถาปัตยกรรมของระบบ (System Architecture Design)
4. ออกแบบฐานข้อมูล (Persistent Data Design)



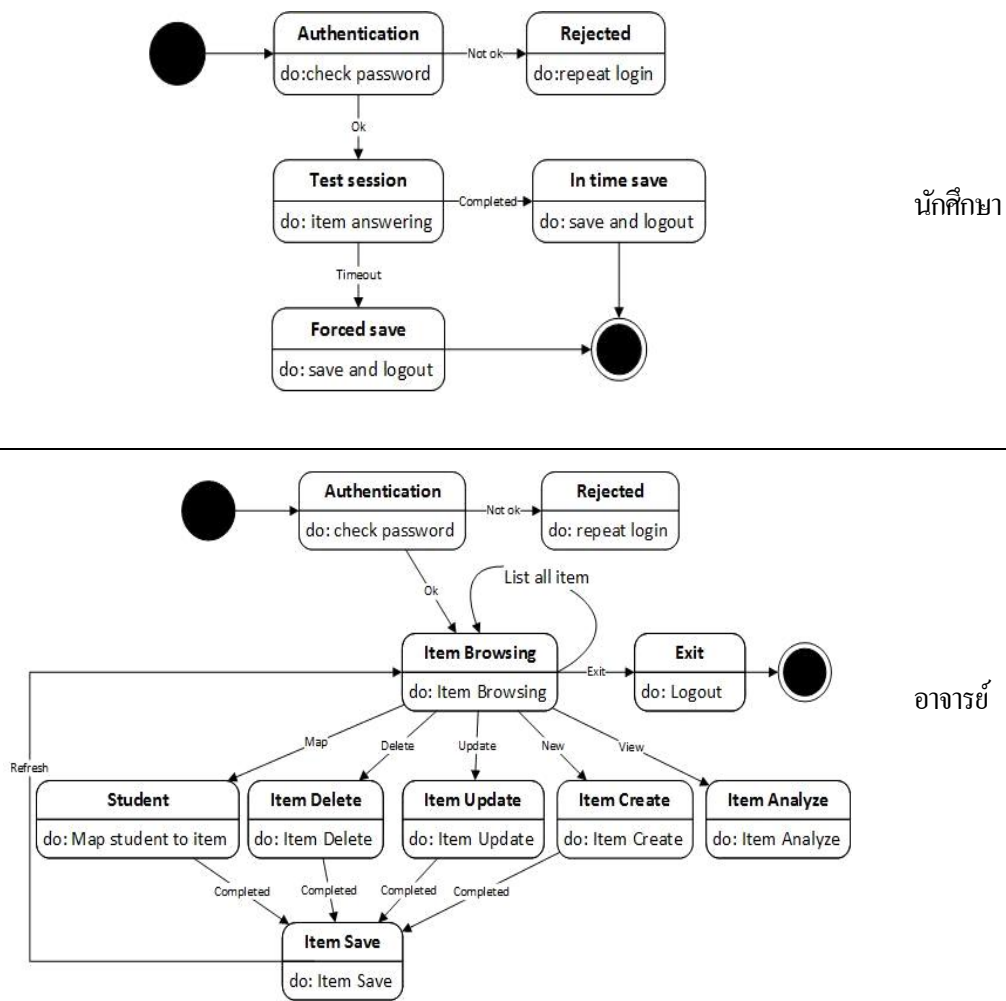
รูปที่ 7 Use Case Diagram



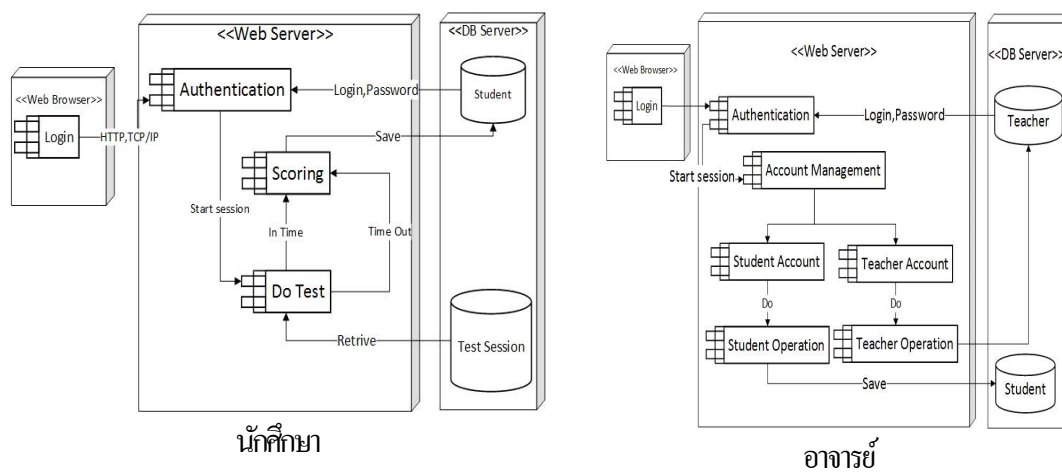
รูปที่ 8 Class diagram



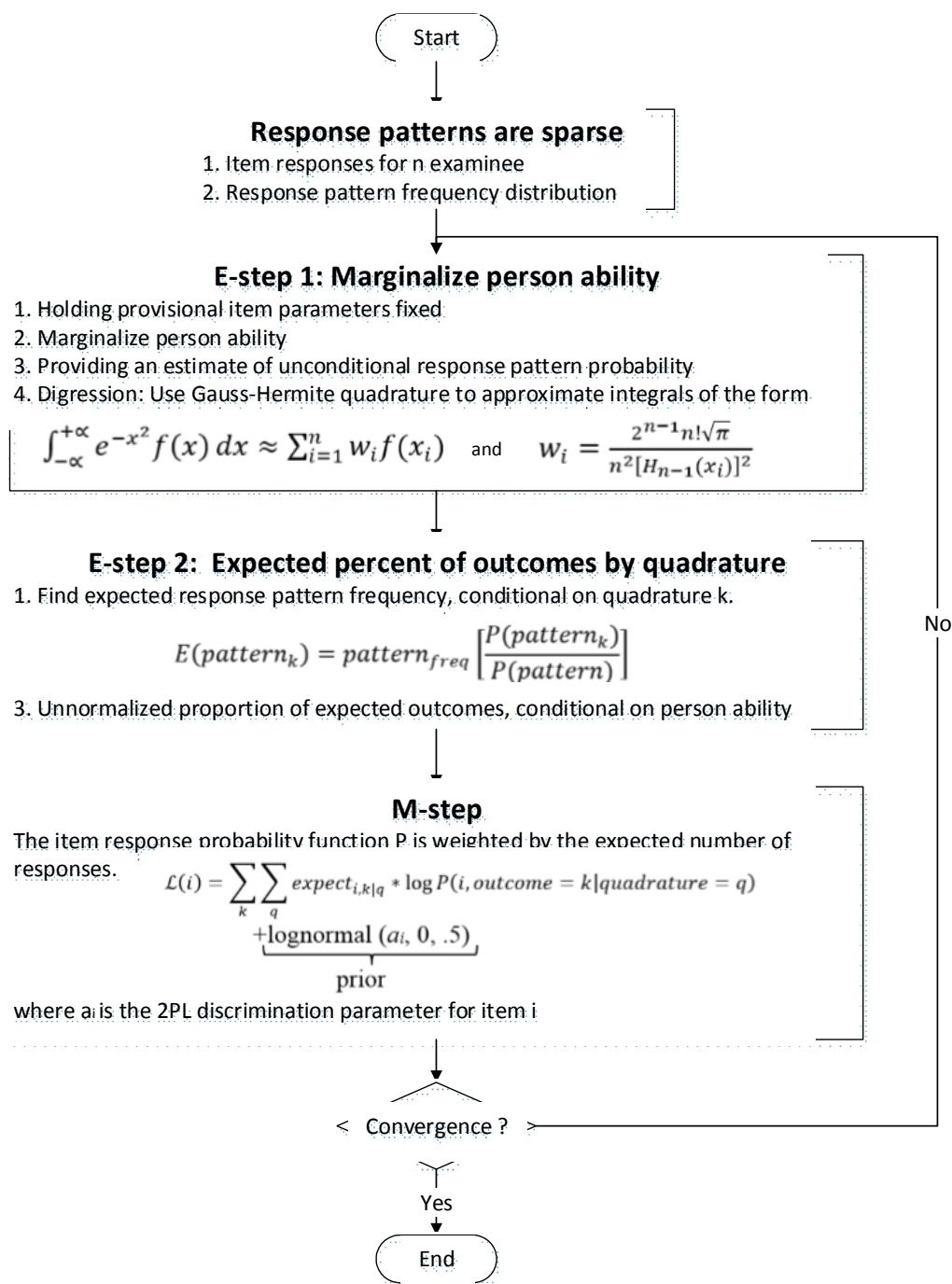
รูปที่ 9 Sequence Diagram



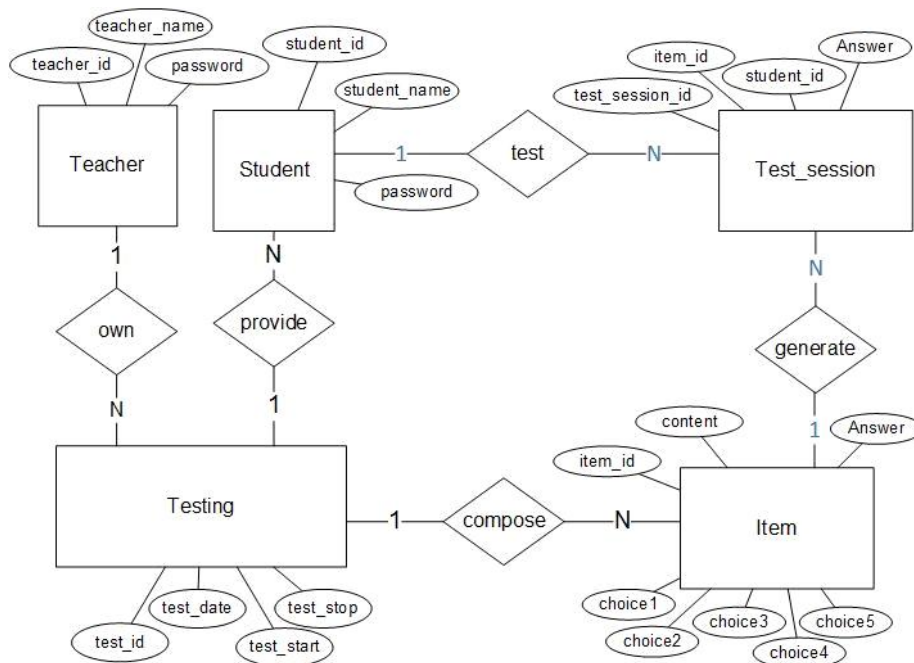
รูปที่ 10 State Diagram



รูปที่ 11 Deployment Diagram



รูปที่ 12 โฟลว์ชาร์ตการประมาณค่าพารามิเตอร์ด้วยอัลกอริทึมการคาดหวัง-การหาจุดสูงสุด



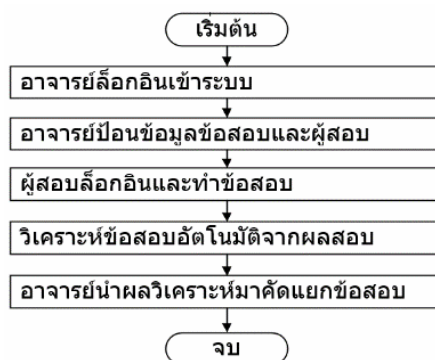
รูปที่ 13 Entity-Relationship Diagram ผลการดำเนินงาน

ผลการดำเนินงาน

หลังจากได้วิเคราะห์และออกแบบระบบแล้ว ขั้นตอนต่อมาคือการพัฒนาระบบซึ่งมีระบบย่อยหลายระบบ เริ่มจากระบบจัดการข้อสอบของอาจารย์ไปจนถึงระบบวิเคราะห์ข้อสอบอัตโนมัติอันเป็นขั้นตอนสุดท้าย

ระบบเริ่มต้นโดยอาจารย์จะล็อกอินเข้าระบบเพื่อจัดการข้อสอบซึ่งมีข้อมูลหลายส่วน ได้แก่ ข้อมูลวันเดือนปีที่จะสอบ ข้อมูลโจทย์คำถาม ข้อมูลตัวเลือก ข้อมูลคำตอบ และส่วนสุดท้ายคือข้อมูลผู้สอบ เมื่อถึงวันสอบผู้สอบจะ

ล็อกอินเข้าระบบมาทำข้อสอบซึ่งโจทย์ข้อสอบของผู้สอบแต่ละคนจะมีการเรียงลำดับคำถาม และตัวเลือกแบบสุ่ม จึงมีความแตกต่างกันเพื่อป้องกันการแอบมองคำตอบกันของผู้สอบและระบบจะตั้งเวลาไว้ถ้าหมดเวลาระบบจะบังคับการบันทึกข้อสอบอัตโนมัติแล้วออกนอกระบบทันที และเมื่อนักศึกษาทำข้อสอบเสร็จแล้ว ระบบจะรวมคะแนนสอบและวิเคราะห์ข้อสอบอัตโนมัติ อาจารย์สามารถคัดแยกข้อสอบจากผลการวิเคราะห์ข้อสอบได้ทันที การทำงานของระบบเป็นไปตามผังโปรแกรมต่อไปนี้



รูปที่ 14 ขั้นตอนการทำงานของระบบการวิเคราะห์ข้อสอบออนไลน์อัตโนมัติ

ข้อ	ข้อ	ความยาก	อำนาจจำแนก	การเดา	
☐	1	-2.06 = ง่ายมาก	1.21 = จำแนกได้ปานกลาง	0.43 = เดาง่ายมาก	
☐	2	-2.06 = ง่ายมาก	1.21 = จำแนกได้ปานกลาง	0.43 = เดาง่ายมาก	
☐	3	-1.92 = ง่าย	1.80 = จำแนกได้สูงมาก	0.14 = เดายาก	
☐	4	-2.06 = ง่ายมาก	1.21 = จำแนกได้ปานกลาง	0.43 = เดาง่ายมาก	
☐	5	-0.59 = ง่าย	0.51 = จำแนกได้ต่ำ	-0.43 = เดายากมาก	
☐	6	-0.59 = ง่าย	0.51 = จำแนกได้ต่ำ	-0.43 = เดายากมาก	
☐	7	-2.06 = ง่ายมาก	1.21 = จำแนกได้ปานกลาง	0.43 = เดาง่ายมาก	
☐	8	-2.57 = ง่ายมาก	0.28 = จำแนกได้ต่ำมาก	0.71 = เดาง่ายมาก	
☐	9	-0.92 = ง่าย	0.62 = จำแนกได้ต่ำ	-0.14 = เดายากมาก	
☐	10	-2.57 = ง่ายมาก	0.28 = จำแนกได้ต่ำมาก	0.71 = เดาง่ายมาก	

รูปที่ 18 ผลการวิเคราะห์ข้อสอบแบบ IRT

วิจารณ์

เนื่องจากกลุ่มตัวอย่างที่นำมาใช้ในการวิจัยมีขนาดเล็กมากทำให้ไม่สามารถเปรียบเทียบความแตกต่างระหว่างผลการวิเคราะห์ข้อสอบด้วยทฤษฎีดั้งเดิมกับทฤษฎีการตอบสนองข้อสอบได้อย่างชัดเจน

สรุปผล

ระบบที่พัฒนาเสร็จแล้วเมื่อนำมาใช้งานจริงได้ค้นพบข้อเท็จจริงบางประการ ได้แก่

1. เวลาโดยเฉลี่ยในการทำข้อสอบแบบออนไลน์จะน้อยกว่าเวลาเฉลี่ยที่ทำข้อสอบด้วยกระดาษแบบดั้งเดิมในจำนวนข้อที่เท่ากัน สาเหตุน่าจะมาจากยูสเซอร์อินเทอร์เฟซแบบใช้เมาส์คลิกที่ใช้เวลาน้อยกว่าใช้มือฝนด้วยดินสอตำหรือกากบาทข้อสอบกระดาษ

2. กระบวนการวิเคราะห์ข้อสอบทั้งสองทฤษฎีล้วนใช้ค่าสถิติของข้อมูลผลสอบมาวิเคราะห์ ทำให้ขนาดของข้อมูลมีส่วนสำคัญต่อผลการวิเคราะห์ ดังนั้นจำนวนผู้เข้าสอบมากจะมีความน่าเชื่อถือในผลการวิเคราะห์มากกว่าจำนวนผู้เข้าสอบน้อย

3. แม้ว่าระบบการสอบออนไลน์จะแก้ปัญหาผู้สอบแอบมองคำตอบกัน เพราะระบบจะสุ่มข้อสอบและตัวเลือกทำให้นักศึกษาแต่ละคนมีการเรียงลำดับโจทย์คำถามและตัวเลือกแตกต่างกัน แต่การเดาข้อสอบถูกของนักศึกษาที่ไม่มีความรู้จริงก็ยังเป็นตัวแปรสำคัญที่จะทำให้ผลการวิเคราะห์คลาดเคลื่อนจากความเป็นจริง

4. กรณีที่นักศึกษาเดาข้อสอบหรือตอบโดยไม่มีความรู้เลยจะมีโอกาสได้คะแนนประมาณ 25% ในกรณีข้อสอบ 4 ตัวเลือกหรือ 1 ใน 4 ดังนั้นกรณีที่นักศึกษาได้คะแนนต่ำกว่า 25% ของคะแนนเต็มแสดงว่านักศึกษาไม่มีความรู้โดยสิ้นเชิง และไม่มีแม้วิจารณ์ฐานในการเดาข้อสอบ ซึ่งจัดว่าเป็นนักศึกษาที่อ่อนมาก

5. การวิเคราะห์ข้อสอบด้วยทฤษฎีการตอบสนองข้อสอบใช้เวลามากกว่าทฤษฎีการสอบแบบดั้งเดิมอย่างมีนัยสำคัญและจะแตกต่างกันมากยิ่งขึ้นถ้ามีจำนวนผู้เข้าสอบและจำนวนข้อสอบมากขึ้นเนื่องจากอัลกอริทึม E-M ที่ใช้ในการประมาณค่าพารามิเตอร์ต้องมีการวนซ้ำรอบการทำงานที่เป็นสัดส่วนโดยตรงต่อจำนวนผู้เข้าสอบที่มีอัตราการขยายขนาดแบบเอ็กซ์โปเนนเชียลด้วยความซับซ้อนเชิงเวลาที่วัดได้แบบ $O(kn)$, $k > 1$

ข้อเสนอแนะจากการวิจัย

1. ระบบการสอบออนไลน์ใช้ระยะเวลาในการทำข้อสอบน้อยกว่าการทำข้อสอบแบบดั้งเดิมด้วยกระดาษ ดังนั้นอาจจะกำหนดช่วงระยะเวลาสอบในแต่ละวิชาให้น้อยลงกว่าการสอบด้วยกระดาษ เช่น จากเดิมสอบ 3 ชั่วโมง อาจลดลงเหลือ 2.45 ชั่วโมง หรือ 3 ชั่วโมงเท่าเดิมแต่เพิ่มจำนวนข้อให้มากขึ้นอีก 10% - 15% เป็นต้น

2. แม้ว่าจะไม่สามารถป้องกันการเดาข้อสอบของผู้สอบที่ก่อให้เกิดปัญหาความคลาดเคลื่อนจากการวิเคราะห์ข้อสอบได้ แต่การวิเคราะห์ข้อสอบด้วยทฤษฎีการตอบสนอง

ข้อสอบสามารถตรวจจับค่าความง่ายต่อการเดาได้ ดังนั้น น่าจะมีการปรับระบบการให้คะแนนใหม่โดยให้น้ำหนักแต่ละข้อไม่เท่ากันตามความยากง่ายต่อการเดาซึ่งจะทำให้ได้คะแนนความสามารถที่แท้จริงในการสอบของนักศึกษา

3. ทฤษฎีการตอบสนองข้อสอบยังมีแนวทางเพื่อวิจัยและประยุกต์ได้อีก เช่น โมเดลการตอบสนองเกรด (Graded Respond Model: GRM) การทดสอบที่ปรับตัวเองได้ด้วยคอมพิวเตอร์ (Computerized Adaptive Testing: CAT) และทฤษฎีการตอบสนองข้อสอบแบบหลายมิติ (Multidimensional Item Respond Theory: MIRT)

4. ในเปรียบเทียบผลการวิเคราะห์ข้อสอบด้วยทฤษฎีดั้งเดิมกับทฤษฎีการตอบสนองข้อสอบให้ได้ผล

ที่น่าเชื่อถือ ควรใช้กลุ่มตัวอย่างตั้งแต่ 200 – 500 คนในการนี้ทฤษฎีดั้งเดิม และ 500 – 1,000 คน ในการนี้ทฤษฎีการตอบสนองข้อสอบ (Royce Hernandez, 2009)

5. การประมาณค่าพารามิเตอร์ที่กินเวลาค่อนข้างมาก เพราะใช้วิธีการมาร์จินัลแมกซ์ลิคิฮูด (Marginal Maximum Likelihood Method) ของ Bock และ Aitkin (1981) ซึ่งจัดเป็นอัลกอริทึมการคาดหวัง-การหาจุดสูงสุด (Expectation-Maximization: E-M) มีประเด็นที่ควรพิจารณาโดยเฉพาะการระบบผ่านเว็บซึ่งมีความหน่วงในการแสดงผลมาก ถ้าต้องรอผลลัพธ์จากการประมวลผลที่นานเพิ่มเข้าไปอีกจะทำให้เหมือนระบบหยุดทำงานชั่วคราว ซึ่งควรปรับปรุงด้วยอัลกอริทึมที่ประมวลผลได้เร็วกว่า มาแทนวิธีการที่ใช้ในการวิจัยนี้

บรรณานุกรม

- กิตติมา เจริญศิริ (2550). "การพัฒนาระบบการวิเคราะห์และการสร้างคลังข้อสอบผ่านเครือข่ายอินเทอร์เน็ตสำหรับมหาวิทยาลัยเอกชน". วารสารศรีปทุมปริทัศน์. 7(2) : หน้า 5-12
- ฉัตรศิริ ปิยะพิมลสิทธิ์ (2548). การวิเคราะห์คุณภาพข้อสอบด้วยโปรแกรมคอมพิวเตอร์. [ออนไลน์]. สืบค้นเมื่อวันที่ 30 เมษายน จาก <http://www.watpon.com/Elearning/itemanalysis.pdf>.
- ศิริชัย กาญจนวาสี. (2555). ทฤษฎีการทดสอบแนวใหม่ (ฉบับปรับปรุง). กรุงเทพฯ. จุฬาลงกรณ์มหาวิทยาลัย.
- สุโกศล วโนทยาพิทักษ์. (2551). "การพัฒนาระบบการสอบออนไลน์ผ่านเครือข่ายอินเทอร์เน็ต". วารสารมหาวิทยาลัยคริสเตียน. 14(2) : หน้า 117-127
- สุโกศล วโนทยาพิทักษ์. (2555). "การพัฒนาระบบคลังข้อสอบออนไลน์แบบมีวงจรการวิเคราะห์ข้อสอบอัตโนมิติ". วารสารวิศวกรรมศาสตร์ มหาวิทยาลัยสยาม. 2(25) : หน้า 22-32
- Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Monterey, CA: Brooks/Cole Publishing.
- AtulKahate. (2004). *Object Oriented Analysis & Design*. Tata McGraw-Hill Education.
- Bock, R.D.; Aitkin, M. (1981). *Marginal maximum likelihood estimation of item parameters : application of an EM algorithm*. *Psychometrika*. 46(4) : 443-459.
- Linn, R. L. (Ed.). (1989). *Educational Measurement: Third Edition*, New York: Macmillan Publishing Co.
- Royce Hernandez. (2009). *Comparison of the Item Discrimination and Item Difficulty of the Quick-Mental Aptitude Test using CTT and IRT Methods*. The International Journal of Educational and Psychological Assessment April 2009, 1(1) : 12-18

- RuiminShen, Yiyang Tang, Tong Zhen Zhang. (2001). *The Intelligent Assessment System in Web-based Distance Learning Education*. 31th ASSEE/IEEE Frontiers in Education Conference: session T1F
- Williams, H.E. (2004) .*Web Database Application with PHP and MySQL: Building Effective Database-driven Web Sites*. O'Reilly Media Inc.
- Yuan Zhenming, Zhang Liang, Zhan Guohua. (2003). *A Novel Web-based Online Examination System for Computer Science Education*. 33th ASSEE/IEEE Frontiers in Education Conference: session S3F, November.

