

จุฬาลงกรณ์มหาวิทยาลัย

ทุนพัฒนาอาจารย์ใหม่/นักวิจัยใหม่

กองทุนรัชดาภิเษกสมโภช

รายงานการวิจัยเรื่อง

การพัฒนาค้นข้อสอบวิชา 2702303

การวัดและการประเมินผลทางการศึกษา

ด้วยโปรแกรมคอมพิวเตอร์

โดย

ศูนย์วิทยทรัพยากร
ณัฐภรณ์ หลาวทอง
จุฬาลงกรณ์มหาวิทยาลัย

มีนาคม 2547

ISBN 974-13-2901-6

กิตติกรรมประกาศ

รายงานการวิจัยเรื่องการพัฒนาคลังข้อสอบวิชา 2702303 การวัดและการประเมินผลทางการศึกษาด้วยโปรแกรมคอมพิวเตอร์ เป็นการวิจัยเชิงพัฒนา (Developmental Research) จุดเริ่มต้นของการวิจัยนี้จึงมีความเกี่ยวข้องกับความพยายามในทำให้กระบวนการคัดเลือกและจัดเก็บข้อสอบที่มีคุณภาพอย่างเป็นระบบ การพิจารณาคัดเลือกข้อสอบแต่ละครั้งอยู่ภายใต้ความเหมาะสมของฐานข้อมูลที่เกี่ยวข้องกับเนื้อหาวิชา พฤติกรรมที่ต้องการวัด และคุณภาพของข้อสอบรายข้อ ได้แก่ ค่าความยาก และอำนาจจำแนก

การวิจัยจะสำเร็จลงได้ยาก หากปราศจากการสร้างข้อสอบที่มีคุณภาพจากคณะครูผู้สอนรายวิชา 2702302 และ 2702303 ทุกท่านที่เคยสอนและรับผิดชอบสอนวิชาดังกล่าวในปัจจุบัน ได้แก่ รศ.ดร.ศิริเดช สุชีวะ ผศ.ดร.อวยพร เรืองตระกูล ผศ.ดร.ดวงกมล ไตรวิจิตรคุณ ผศ.ดร.เอมอร จังศิริพรภรณ์ ผศ.ดร.สุพิมพ์ ศรีพันธ์วรสกุล อ.ดร.วรรณิ์ แกมเกตุ และผู้วิจัยเอง ดังนั้นจึงขอขอบคุณคณาจารย์ผู้สอนเป็นอย่างสูง โดยเฉพาะอย่างยิ่ง ผศ.ดร.ดวงกมล ไตรวิจิตรคุณ ผู้เก็บรวบรวมเอกสารข้อสอบที่ผ่านมาอย่างเป็นระบบ รวมทั้งการวิเคราะห์คุณภาพข้อสอบในปีการศึกษา 2541 ซึ่งเป็นข้อมูลอันเป็นประโยชน์และสำคัญอย่างยิ่งต่อการดำเนินการวิจัยนี้

การศึกษาและพัฒนาโปรแกรมคอมพิวเตอร์เป็นเรื่องที่ต้องใช้เวลาและความเชี่ยวชาญเฉพาะทาง ผู้วิจัยขอขอบคุณ นายพนัส จันท์เปล่ง ที่ร่วมกำหนดแนวทางโครงสร้างคลังข้อสอบวิชา 2702303 ออกมาให้เป็นรูปธรรมสามารถนำไปใช้ได้จริง นางสาวธิดาพร โดสดี และนายพงษ์สิทธิ์ เพชรผล ที่ช่วยบันทึกข้อมูลการวิจัย ผู้วิจัยหวังเป็นอย่างยิ่งว่างานวิจัยนี้นอกจากจะเป็นประโยชน์ต่อคณะครูผู้สอน และช่วยลดภาระงานของ คุณอุทุมพร ชาติเผือก ผู้เป็นกำลังสำคัญในการจัดพิมพ์ข้อสอบเสมอมา

งานวิจัยนี้ผู้วิจัยได้รับความกรุณาเป็นอย่างสูงตลอดมาจาก รศ.ดร.ศิริชัย กาญจนวาสี อาจารย์อาวุโส ซึ่งท่านได้ให้ข้อเสนอแนะอันเป็นประโยชน์อย่างยิ่งตั้งแต่เริ่มเขียนโครงร่างการวิจัยจนกระทั่งงานวิจัยเสร็จสมบูรณ์

ผู้วิจัยขอกราบขอบพระคุณเป็นอย่างสูง ในความกรุณาของคณะกรรมการพิจารณาอบทุนพัฒนาอาจารย์ใหม่/นักวิจัยใหม่ กองทุนรัชดาภิเษกสมโภช จุฬาลงกรณ์มหาวิทยาลัย เพื่อให้ในการสนับสนุนด้านงบประมาณของในการดำเนินการวิจัย การเดินทางไปนำเสนอผลการวิจัยและการฝึกอบรมเกี่ยวกับการวิจัยสถาบัน ณ มลรัฐฟลอริดา ประเทศสหรัฐอเมริกา ผู้วิจัยได้ใช้เงินทุนของท่านเพื่อประโยชน์ในการพัฒนาความรู้ความสามารถทางวิชาการอย่างแท้จริง

ณัฐภรณ์ หลาวทอง

มีนาคม 2547

บทคัดย่อ

เรื่อง การพัฒนาคลังข้อสอบวิชา 2702303 การวัดและการประเมินผลทางการศึกษา
ด้วยโปรแกรมคอมพิวเตอร์
(The Development of Item Bank for 2702303 Educational Measurement and
Evaluation Course by Computer Program)

ผู้วิจัย ณัฐภรณ์ หลาวทอง

การวิจัยครั้งนี้มีวัตถุประสงค์ เพื่อศึกษาและจำแนกคุณลักษณะของข้อสอบวิชา 2702303 การวัดและการประเมินผลทางการศึกษา ตามเกณฑ์พฤติกรรมการเรียนรู้ ประเภทของเนื้อหาวิชาที่มุ่งวัด และคุณภาพข้อสอบรายข้อด้านความยาก (p) และอำนาจจำแนก (r) และเพื่อพัฒนาคลังข้อสอบ โดยใช้โปรแกรมคอมพิวเตอร์ สำหรับเก็บรวบรวมข้อสอบที่มีคุณภาพไว้ใช้ในโอกาสต่อไป ข้อสอบที่นำมาศึกษาประกอบด้วย ข้อสอบวิชา 2702303 การวัดและการประเมินผลทางการศึกษา และข้อสอบในวิชา 2702302 การประเมินผลการเรียนการสอน จำนวน 597 ข้อ วิเคราะห์ข้อมูลด้วยสถิติบรรยาย เช่น ความถี่ ร้อยละ ค่าเฉลี่ยเลขคณิต ส่วนเบี่ยงเบนมาตรฐาน ค่าความเบ้ ค่าความโด่ง ค่าความยาก และค่าอำนาจจำแนก

ผลการวิจัย มีดังนี้

1. ข้อสอบวัดพฤติกรรมการเรียนรู้เรียงลำดับดังนี้ ความเข้าใจ (31%) ความรู้ความจำ (28.1%) การวิเคราะห์ (26%) การนำไปใช้ (10.1%) การสังเคราะห์ (3%) การประเมิน (1.8%) เนื้อหาวิชาที่มุ่งวัดมีจำนวนข้อสอบมากที่สุด คือ หลักการวัดประเมินผล การวัดประเมินผลตามพระราชบัญญัติการศึกษา 2542 (15.9%) รองลงมาคือเรื่องการตรวจสอบคุณภาพเครื่องมือทั้งฉบับ (13.7%) ข้อสอบมีคุณภาพผ่านเกณฑ์ค่าความยากและอำนาจจำแนกมีจำนวนทั้งสิ้น 262 ข้อ (43.89%) ค่าความยากเฉลี่ยเท่ากับ 0.63 ค่าอำนาจจำแนกเฉลี่ยเท่ากับ 0.26 ค่าความเที่ยงของแบบสอบมีค่าระหว่าง 0.72-0.81

2. การพัฒนาคลังข้อสอบวิชา 2702303 การวัดและการประเมินผลทางการศึกษา ด้วยโปรแกรม ไมโครซอฟต์แอสเซส ทำหน้าที่จัดเก็บข้อสอบที่คัดเลือกแล้วไว้ในคลังข้อสอบ สำหรับสร้างแบบสอบที่จะนำไปใช้ต่อไป โดยข้อสอบที่จัดเก็บไว้จำแนกตาม ปีการศึกษา ภาคการศึกษา ความยาก อำนาจจำแนก ระดับพฤติกรรมการเรียนรู้ และเนื้อหา

Abstract

Title The Development of Item Bank for 2702303 Educational Measurement and Evaluation Course by Computer Program

Researcher Nuttaporn Lawthong

The objectives of this research were to classify test items for 2702303 Educational Measurement and Evaluation course based on Bloom's taxonomy of Knowledge, to find their difficulty and discrimination indices and to write an item bank program. The items were from two courses: 2702302 Evaluation of Learning and Teaching, and 2702303 Educational Measurement and Evaluation. 597 items were analyzed by descriptive statistics.

The research findings were as follows:

1) There were more items to measure comprehension (31%) than other levels, viz. knowledge (28.1%), analysis (26%), application (10.1%), synthesis (3%), and evaluation (1.8%). There were more items to measure principles of measurement and evaluation (15.9%) than other topics such as validity and reliability (13.7%). 262 items were qualified to put into the item bank (43.89%). The average of difficulty indices was 0.63 and that of discrimination was 0.26. The reliability indices of 6 tests were ranging from 0.72 to 0.81.

2) The development of an Item Bank for Educational Measurement and Evaluation Course 2702303 by Microsoft Access program played a role in collecting selected items into the bank for later test building. Year, semester, difficulty index, discrimination index, Bloom's taxonomy of Knowledge, and topic classified used items.

สารบัญเนื้อหา

		หน้า
	กิตติกรรมประกาศ	๗
	บทคัดย่อ	๘
	สารบัญเนื้อหา	๙
	สารบัญตาราง	๗
	สารบัญภาพ	๗
บทที่ 1	บทนำ	1
	ความเป็นมาและความสำคัญของปัญหา	1
	วัตถุประสงค์ของการวิจัย	3
	ขอบเขตของการวิจัย	3
	ข้อตกลงเบื้องต้น	4
	ข้อจำกัดในการวิจัย	4
	คำจำกัดความที่ใช้ในงานวิจัย	4
	ประโยชน์ที่คาดว่าจะได้รับ	5
บทที่ 2	เอกสารและงานวิจัยที่เกี่ยวข้อง	6
	ตอนที่ 1 รายละเอียดเกี่ยวกับ วิชา 2702303	
	การวัดและการประเมินผลทางการศึกษา	6
	ตอนที่ 2 การพิจารณาจำแนกข้อสอบ	8
	ตอนที่ 3 คลังข้อสอบ	15
บทที่ 3	วิธีดำเนินการวิจัย	36
	ข้อสอบที่นำมาวิเคราะห์	36
	การพัฒนาโปรแกรมคอมพิวเตอร์	36
	การเก็บรวบรวมข้อมูล	37
	การวิเคราะห์ข้อมูล	38
บทที่ 4	ผลการวิเคราะห์ข้อมูล	39
	ตอนที่ 1 ผลการวิเคราะห์ข้อสอบวิชา 2702303	
	การวัดและการประเมินผลการศึกษา	39
	ตอนที่ 2 ผลการวิเคราะห์ข้อสอบที่จะคัดเลือกไว้ในคลังข้อสอบ	47
	ตอนที่ 3 ผลการวิเคราะห์ค่าความเที่ยงของแบบสอบ	51

		หน้า
	ตอนที่ 4 ลักษณะคลังข้อสอบวิชา2702303	
	การวัดและการประเมินผลการศึกษา	52
บทที่ 5	สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ	62
	สรุปผลการวิจัย	62
	อภิปรายผล	65
	ข้อเสนอแนะ	67
บรรณานุกรม		69
ภาคผนวก		71



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญตาราง

ตารางที่		หน้า
1	ความถี่ร้อยละของข้อสอบจำแนกตามตัวแปร	40
2	ความถี่และร้อยละ จำแนกตามเนื้อหาวิชาและพฤติกรรมการเรียนรู้	41
3	ค่าเฉลี่ยเลขคณิต ส่วนเบี่ยงเบนมาตรฐาน ความแปร และความโค้งของ ค่าความยากและอำนาจจำแนกของข้อสอบ	43
4	ค่าความถี่และร้อยละของคุณภาพข้อสอบจำแนกตามความยากและ การนำไปใช้	43
5	ค่าความถี่และร้อยละของคุณภาพข้อสอบจำแนกตามความยากและ การนำไปใช้	44
6	ความถี่และร้อยละของข้อสอบที่มีคุณภาพด้านความยากและอำนาจจำแนก	44
7	ความถี่และร้อยละของคุณภาพข้อสอบจำแนกตามความยากและ อำนาจจำแนก	45
8	ความถี่ร้อยละของข้อสอบที่สามารถเก็บไว้ใช้ได้ จำแนกตามตัวแปร	48
9	ความถี่และร้อยละข้อสอบที่เก็บไว้ใช้ได้จำแนกตามเนื้อหาวิชาและ พฤติกรรมการเรียนรู้	49
10	ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ความเบี่ยงเบนแบบ KR-20และ ความคลาดเคลื่อนมาตรฐาน จำแนกตามปีการศึกษาและภาคการเรียน	51

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญภาพ

		หน้า
ภาพที่ 1	การกระจายค่าความยากและอำนาจจำแนกของข้อสอบ	46
ภาพที่ 2	หน้าจอโปรแกรมคลังข้อสอบ	52
ภาพที่ 3	การจัดเก็บข้อสอบ	53
ภาพที่ 4	การค้นหาตามพฤติกรรมการเรียนรู้	54
ภาพที่ 5	การค้นหาตามเนื้อหาวิชา	55
ภาพที่ 6	การค้นหาตามความยาก	56
ภาพที่ 7	การค้นหาตามอำนาจจำแนก	57
ภาพที่ 8	ข้อมูลที่ได้จากการค้นหา	58
ภาพที่ 9	การจัดพิมพ์ในโปรแกรมคลังข้อสอบ	59
ภาพที่ 10	เพิ่มข้อมูลที่จัดเก็บในโปรแกรม Microsoft Word	60
ภาพที่ 11	กระบวนการใช้คลังข้อสอบวิชาการวัดและการประเมินผลทางการศึกษา	61



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

หลักสูตรครุศาสตรบัณฑิต (หลักสูตรปรับปรุง พ.ศ. 2543) กำหนดให้วิชา 2702303 การวัดและการประเมินผลทางการศึกษา จัดอยู่ในหมวดวิชาครู เป็นวิชาบังคับ 2 หน่วยกิต จัดการเรียนการสอนโดยภาควิชาวิจัยการศึกษา คณะครุศาสตร์ เปิดสอนทั้งสองภาคการศึกษา ตั้งแต่ปีการศึกษา 2545 โดยผู้สอน 5-6 ท่าน รับผิดชอบจัดการเรียนการสอนแยกตามตอนเรียน เพื่อให้เกิดความเป็นมาตรฐานในการวัดประเมินผล ครูผู้สอนจะกำหนดน้ำหนักและตั้งเกณฑ์การประเมินร่วมกัน ในการออกข้อสอบกลางภาคและปลายภาค ผู้สอนแต่ละท่านจะได้รับมอบหมายให้ออกข้อสอบตามตารางแสดงแผนผังข้อสอบ (Testing blueprint) แล้วประชุมร่วมกันเพื่อคัดเลือกข้อสอบอีกครั้ง โดยในแต่ละครั้งของการออกข้อสอบจะมีการคัดเลือกข้อสอบที่มีการสร้างมาแล้วในปีการศึกษาที่ผ่านมา เพื่อนำมาใช้ในแบบสอบฉบับที่สร้างขึ้นใหม่ และในการสอบแต่ละครั้งมีทั้งแบบสอบประเภทอัตนัยและปรนัย

เพื่อให้การจัดเก็บข้อสอบที่มีคุณภาพเป็นไปอย่างมีระบบ ผู้วิจัยมีความสนใจพัฒนาคคลังข้อสอบวิชาการวัดและการประเมินผลทางการศึกษาด้วยโปรแกรมคอมพิวเตอร์ สำหรับจัดเก็บข้อสอบที่มีคุณภาพ สำหรับฐานข้อมูลข้อสอบประกอบด้วย ข้อสอบในวิชา 2702303 การวัดและการประเมินผลทางการศึกษา ปีการศึกษา 2545 และข้อสอบในวิชา 2702302 การประเมินผลการเรียนการสอน ปีการศึกษา 2540-2544 ซึ่งเป็นวิชาในหลักสูตรเดิมที่มีเนื้อหาบางส่วนใกล้เคียงกันและได้มีการจัดเก็บแบบสอบต้นฉบับไว้แล้ว โดยผู้วิจัยจะคัดเลือกเฉพาะข้อสอบที่มีเนื้อหาที่สอดคล้องกับวิชา 2702303 มาเป็นข้อมูลในการศึกษา

การพัฒนาคคลังข้อสอบวิชา 2702303 การวัดและการประเมินผลทางการศึกษา ด้วยโปรแกรมคอมพิวเตอร์จะช่วยให้การจัดเก็บข้อสอบเป็นไปอย่างมีระบบ ทำให้คัดเลือกข้อสอบได้ตรงตามวัตถุประสงค์ / เนื้อหา พร้อมทั้งได้ข้อสอบที่มีคุณภาพทั้งด้านความยากและอำนาจจำแนก สามารถเตรียมข้อสอบคู่ขนานได้ง่าย มีความสะดวกมากยิ่งขึ้น และช่วยให้ประหยัดเวลาออกข้อสอบอีกด้วย สอดคล้องกับคำกล่าวของ ศิริชัย กาญจนวาสิ (2544) เกี่ยวกับ คลังข้อสอบว่าเป็นที่รวมข้อสอบที่มีคุณภาพจำนวนมากสำหรับใช้วัดความรู้ ความสามารถ และทักษะอันเป็นผล การเรียนรู้ของผู้เรียนตามจุดมุ่งหมายของวิชาและหลักสูตร ปัจจุบันสามารถใช้ไมโครคอมพิวเตอร์ช่วยเก็บข้อสอบไว้เป็นคลังข้อสอบ ผู้สอนสามารถเลือกหรือสุ่มข้อสอบจากคลังข้อสอบให้ครอบคลุมจุดมุ่งหมาย และเนื้อหาที่ต้องการสอบในแต่ละครั้งของการสอบ อันจะเป็นประโยชน์ต่อการ

พัฒนาแบบสอบคู่ขนาน การปรับเทียบคะแนนระหว่างแบบสอบและใช้คอมพิวเตอร์พิมพ์ข้อสอบ เพื่อนำมาใช้เป็นแบบสอบสำหรับวัดผลการเรียนรู้ของผู้เรียนได้อย่างง่าย สะดวก และมีประสิทธิภาพ

จากการศึกษางานวิจัยเกี่ยวกับการพัฒนาค้างข้อสอบที่ผ่านมา 6 ฉบับ พบว่ามีประเด็น การพิจารณา 3 ประการ คือ ประเด็นแรก หน้าที่การทำงานของค้างข้อสอบ ประเด็นที่สอง ลักษณะแบบสอบ และ ประเด็นสุดท้าย ระบบปฏิบัติการโปรแกรมคอมพิวเตอร์ที่ใช้ในการสร้าง ค้างข้อสอบ

ในประเด็นแรก หน้าที่การทำงานของค้างข้อสอบ จากงานวิจัยที่ศึกษา สามารถแบ่ง ลักษณะค้างข้อสอบออกเป็น 3 ประเภท ได้แก่ การทำหน้าที่ด้านการจัดเก็บข้อสอบ 3 งานวิจัย (ประคอง กรรณสูต, 2536; พลกร กรพิทักษ์, 2533; วีรยุทธ์ ธานี, 2546) การทำหน้าที่จัดเก็บและ วิเคราะห์ข้อสอบ จำนวน 2 งานวิจัย (กฤษฎาพันธ์ พงษ์บริบูรณ์, 2543; วินุลาศ เจริญชัย, 2537) และกลุ่มสุดท้าย คือ การจัดเก็บ ดำเนินการสอบ และวิเคราะห์ข้อสอบ (กอบกิจ สหัสรังษี, 2539)

ประเด็นที่สอง ลักษณะแบบสอบ จากงานวิจัยที่ศึกษา สามารถแบ่งลักษณะแบบสอบ 2 กลุ่ม ได้แก่ ค้างข้อสอบที่จัดเก็บแบบสอบเลือกตอบเพียงอย่างเดียว 5 งานวิจัย (กฤษฎาพันธ์ พงษ์บริบูรณ์, 2543; ประคอง กรรณสูต, 2536; พลกร กรพิทักษ์, 2533; วินุลาศ เจริญชัย, 2537; วีรยุทธ์ ธานี, 2546) โดยงานวิจัยในกลุ่มนี้มีงานวิจัยของ งานวิจัยของ วีรยุทธ์ ธานี ที่สามารถจัด เก็บข้อสอบได้มากกว่าข้อความ โดยสามารถจัดเก็บข้อสอบที่เป็นรูปภาพ หรือสัญลักษณ์ทาง คณิตศาสตร์ได้เพิ่มเติมกว่างานวิจัยทั่วไป สำหรับงานวิจัยของ กฤษฎาพันธ์ พงษ์บริบูรณ์ มีจุดเด่น ในการดำเนินการสอบปฏิบัติการในคอมพิวเตอร์ได้อีกด้วย สำหรับงานวิจัยกลุ่มที่ 2 เป็นของ สามารถจัดเก็บข้อสอบ ได้ทั้ง แบบสอบเลือกตอบ แบบสอบถูกผิด แบบสอบจับคู่ และแบบสอบ เติมคำ (กอบกิจ สหัสรังษี, 2539)

ประเด็นสุดท้าย ระบบปฏิบัติการโปรแกรมคอมพิวเตอร์ที่ใช้ในการสร้างค้างข้อสอบ แบ่ง ออกเป็น 2 กลุ่มคือ การใช้ระบบปฏิบัติการแบบ Dos จำนวน 5 งานวิจัย ภาษาที่นิยมใช้เขียนจะ เป็นภาษา FoxPro ซึ่งจากงานวิจัยที่ผ่านมา สามารถเขียนภาษา ฟอ์แทรน Dbase, Foxbase ได้ อีกด้วย (กฤษฎาพันธ์ พงษ์บริบูรณ์, 2543; กอบกิจ สหัสรังษี, 2539; ประคอง กรรณสูต, 2536; พลกร กรพิทักษ์, 2533; วินุลาศ เจริญชัย, 2537) มีงานวิจัยของ วีรยุทธ์ ธานี (2546) เพียงหนึ่ง เรื่องที่ใช้ระบบปฏิบัติการ Windows พัฒนาค้างข้อสอบ ซึ่งโปรแกรม Microsoft windows มี โปรแกรมเฉพาะการใช้งานให้เลือกสรรมากมาย เช่น Microsoft word, Microsoft powerpoint, Microsoft excel และ Microsoft access ซึ่งโปรแกรมทำยนี้มีความเหมาะสมในการจัดทำฐาน ข้อมูล และยังไม่มีพบว่าม้งานวิจัยที่จัดทำค้างข้อสอบด้วยโปรแกรม Microsoft access ทั้งที่เป็น โปรแกรมที่มีอยู่แล้วสำหรับคอมพิวเตอร์ที่ใช้ระบบปฏิบัติการ Microsoft windows ซึ่งใช้แพร่หลาย

ทั่วโลก ผู้วิจัยจึงมีความสนใจพัฒนาคคลังข้อสอบด้วยโปรแกรม Microsoft access ครั้งนี้ขึ้นใช้ในการจัดเก็บคลังข้อสอบวิชาการวัดและการประเมินผลทางการศึกษา

ในอดีตการจัดเก็บข้อสอบรายวิชา 2702303 ขาดการวางแผนที่เป็นระบบ เนื่องจากมีการจัดเก็บข้อสอบเป็นฉบับๆ ตามปีการศึกษา โดยไม่ได้จำแนกข้อสอบตามวัตถุประสงค์หรือเนื้อหาวิชา รวมถึงพฤติกรรมการเรียนรู้ ข้อสอบบางฉบับมีการวิเคราะห์คุณภาพความยากและอำนาจจำแนกไว้ ด้วยโปรแกรมวิเคราะห์ข้อสอบ (Classical test item analysis) พัฒนาโดย สุพัฒน์ สุกมลสันต์ (2535) แต่บางฉบับก็ไม่ได้วิเคราะห์ ทำให้การเลือกข้อสอบที่มีคุณภาพมาใช้สร้างแบบสอบในปีการศึกษาต่อไปเป็นไปได้ยาก เนื่องจากขาดข้อมูลคุณภาพข้อสอบรายข้อ ดังนั้นแบบสอบที่สร้างในแต่ละปีการศึกษาจึงเป็นการเวียนใช้ข้อสอบเก่า โดยขาดการพิจารณาถึงคุณภาพข้อสอบด้านความยาก และอำนาจจำแนก

ผู้วิจัยเห็นว่าข้อสอบที่จัดเก็บไว้เป็นเวลาประมาณ 7 ปี (ตั้งแต่ปี 2540-2546) ยังขาดระบบที่ดี จึงเห็นว่าควรมีการพัฒนาคลังข้อสอบเพื่อใช้ในวิชา 2702303 การวัดและการประเมินผลทางการศึกษา เพื่อให้การจัดเก็บข้อสอบเป็นระบบและมีการดำเนินการที่ต่อเนื่องมากยิ่งขึ้น เพื่อเป็นประโยชน์ต่อการตัดสินใจคัดเลือกข้อสอบที่มีคุณภาพไปใช้ในปีการศึกษาต่อไปได้ตรงตามวัตถุประสงค์

วัตถุประสงค์การวิจัย

1. เพื่อศึกษาและจำแนกคุณลักษณะของข้อสอบวิชา 2702303 การวัดและการประเมินผลทางการศึกษาตามเกณฑ์พฤติกรรมการเรียนรู้ ประเภทของเนื้อหาวิชา และคุณภาพข้อสอบรายข้อ ด้านความยากและอำนาจจำแนก
2. เพื่อพัฒนาคคลังข้อสอบวิชา 2702303 การวัดและการประเมินผลทางการศึกษาโดยโปรแกรมคอมพิวเตอร์ สำหรับเก็บรวบรวมข้อสอบที่มีคุณภาพไว้ใช้สร้างแบบสอบเพื่อวัดและประเมินผลต่อไป

ขอบเขตของการวิจัย

ข้อสอบที่นำมาศึกษาเป็นข้อสอบปรนัยแบบเลือกตอบประกอบด้วย ข้อสอบในวิชา 2702303 การประเมินผลทางการศึกษา ปีการศึกษา 2545-2546 จำนวน 260 ข้อ และข้อสอบในวิชา 2702302 การประเมินผลการเรียนการสอน ที่มีเนื้อหาสอดคล้องกับรายวิชา 2702303 และได้ผ่านการวิเคราะห์คุณภาพข้อสอบด้านค่าความยากและอำนาจจำแนกแล้ว ประกอบด้วยข้อสอบใน ปีการศึกษา 2541 และ 2544 จำนวน 337 ข้อ รวมข้อสอบทั้งสิ้นจำนวน 597 ข้อ

ข้อตกลงเบื้องต้น

ในการจัดทำข้อสอบแต่ละปีการศึกษา คณะครูผู้สอนรายวิชา 2702302 และ 2702303 ได้มีส่วนร่วมในการออกข้อสอบและประชุมพิจารณาคัดเลือกข้อสอบร่วมกัน ซึ่งถือได้ว่าข้อสอบที่นำมาวิเคราะห์ได้ผ่านการวิเคราะห์ความตรงตามเนื้อหาจากผู้สอนเรียบร้อยแล้ว

ข้อจำกัดในการวิจัย

ข้อสอบที่นำมาวิเคราะห์เป็นข้อสอบปรนัยแบบเลือกตอบเท่านั้น บางข้อมีการนำมาใช้ซ้ำในบางปีการศึกษา แต่มีค่าความยากและอำนาจจำแนกไม่เท่ากันของการนำข้อสอบไปใช้แต่ละครั้ง ซึ่งเป็นผลจากการวิเคราะห์คุณภาพข้อสอบของทฤษฎีการวิเคราะห์แบบดั้งเดิม นอกจากนี้มีข้อสอบบางประเภทที่เป็นชุดข้อสอบที่ใช้สถานการณ์เดียวกัน เพื่อตอบคำถาม 3-4 ข้อ ข้อสอบในลักษณะดังกล่าวนี้จะมีปัญหาซึ่งผู้วิจัยต้องปรับเปลี่ยนข้อคำถามเพื่อการบรรจุลงในคลังข้อสอบที่พัฒนาขึ้น ซึ่งมีลักษณะข้อคำถามเดี่ยวที่อิสระจากคำตอบจากข้อสอบข้ออื่น

เนื่องจากโปรแกรมคลังข้อสอบที่พัฒนาขึ้นจัดเก็บได้เฉพาะข้อที่มีโจทย์คำถามเป็นข้อความเท่านั้น ยังไม่สามารถจัดเก็บข้อสอบที่มี ตาราง สัญลักษณ์ทางสถิติ และรูปภาพประกอบได้ ดังนั้นข้อสอบที่สามารถเก็บข้อสอบได้จึงมีจำนวน 262 ข้อ จากทั้งหมด 322 ข้อ ซึ่งต้องพัฒนาโปรแกรมต่อไป

คำจำกัดความที่ใช้ในการวิจัย

ข้อสอบ หมายถึง ข้อสอบวิชา 2702303 การวัดและการประเมินผลทางการศึกษา ปีการศึกษา 2545 และข้อสอบในวิชา 2702302 การประเมินผลการเรียนการสอน ปีการศึกษา 2540-2544

คลังข้อสอบ หมายถึง เป็นที่รวบรวมและจัดเก็บข้อสอบวิชา 2702303 ที่มีคุณภาพ โดยมีการจัดจำแนกข้อสอบตามเนื้อหาวิชา พฤติกรรมการเรียนรู้ และคุณภาพข้อสอบด้านความยากและอำนาจจำแนก ซึ่งบรรจุไว้ในโปรแกรมคอมพิวเตอร์ Microsoft access

ค่าความยาก หมายถึง สัดส่วนของผู้ตอบข้อสอบนั้นได้ถูกต้องต่อจำนวนผู้สอบเข้าสอบทั้งหมด ซึ่งเป็นการวิเคราะห์ตามทฤษฎีการทดสอบแบบดั้งเดิม มีค่าตั้งแต่ 0 ถึง 1

ค่าอำนาจจำแนก หมายถึง ความสามารถของข้อสอบในการจำแนก แบ่งแยก ความแตกต่างระหว่างผู้สอบที่มีผลสัมฤทธิ์ทางการเรียนหรือทำคะแนนในการสอบต่างกัน ซึ่งเป็นการวิเคราะห์ตามทฤษฎีการทดสอบแบบดั้งเดิม มีค่าตั้งแต่ -1 ถึง 1

คุณภาพข้อสอบ หมายถึง คุณค่าของข้อสอบพิจารณาจากค่าความยาก และอำนาจจำแนกจาก ทฤษฎีการทดสอบแบบดั้งเดิม โดยข้อสอบที่มีคุณภาพจะมีค่าอยู่ระหว่าง 0.2-0.8 และมีค่าอำนาจ จำแนกเท่ากับ 0.2 ขึ้นไป

พฤติกรรมการเรียนรู้ หมายถึง พฤติกรรมการเรียนรู้ด้านพุทธิพิสัย ของ Bloom และคณะ (1956) ประกอบด้วย ความจำ ความเข้าใจ การนำไปใช้ การวิเคราะห์ การสังเคราะห์ และ การประเมินค่า

เนื้อหาวิชา หมายถึง ข้อความรู้เกี่ยวกับ ความสำคัญและความสัมพันธ์ระหว่างวัตถุประสงค์ การศึกษา ประสบการณ์การเรียนรู้ กับกระบวนการวัดและการประเมินผล หลักการวัดและ การประเมินผลการเรียนการสอนตามทฤษฎีดั้งเดิม และทฤษฎีการตอบสนองข้อสอบ ประเภท ลักษณะ และการสร้างเครื่องมือการวัดและการประเมินผลทางการศึกษา เทคนิคการวิเคราะห์ แบบสอบ การสร้างธนาคารข้อสอบ การให้อันดับคะแนน คะแนนมาตรฐาน ปกติวิสัยและ การแปลงความหมายคะแนน หลักและวิธีการประเมินผลภายในและภายนอก เน้นระเบียบ แนวปฏิบัติ และการประยุกต์ในการวัดและการประเมินผลทางการศึกษา

ประโยชน์ที่คาดว่าจะได้รับ

1. ผลจากการพัฒนาคลังข้อสอบวิชา 2702303 การวัดและการประเมินผลทางการศึกษา ทำให้ได้คลังข้อสอบที่บรรจุข้อสอบที่มีคุณภาพ สามารถคัดเลือกนำมาใช้วัดและประเมินผล การศึกษาได้อย่างมีประสิทธิภาพ สะดวกรวดเร็ว
2. เป็นแนวทางในการสร้างคลังข้อสอบในรายวิชาอื่น ๆ ต่อไป

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

เอกสารและงานวิจัยที่เกี่ยวข้องกับการพัฒนาค้างข้อสอบวิชา 2702303 การวัดและการประเมินผลทางการศึกษา ด้วยโปรแกรมคอมพิวเตอร์นี้ แบ่งออกเป็น 3 ตอน ประกอบด้วย

- ตอนที่ 1 รายละเอียดเกี่ยวกับวิชา 2702303 การวัดและการประเมินผลทางการศึกษา
- ตอนที่ 2 การพิจารณาจำแนกข้อสอบ
- ตอนที่ 3 คลังข้อสอบ

ตอนที่ 1 รายละเอียดเกี่ยวกับ วิชา 2702303 การวัดและการประเมินผลทางการศึกษา

1.1 รายละเอียดเกี่ยวกับวิชา

วิชา 2702303 การวัดและการประเมินผลทางการศึกษา (ED MEAS EVA) จำนวน 2 หน่วยกิต บรรยาย 2 ชั่วโมงต่อสัปดาห์ เป็นวิชาบังคับ หลักสูตรครุศาสตรบัณฑิต สำหรับนิสิตระดับปริญญาตรี จัดการเรียนการสอนโดยภาควิชาวิจัยการศึกษา คณะครุศาสตร์ เปิดสอนในภาคการศึกษาต้น ตั้งแต่ปีการศึกษา 2545 พ.ศ. โดยผู้สอน 5-6 ท่าน แยกตามกลุ่มตอนเรียน วิชาที่ต้องเรียนมาก่อนคือ 1. ระเบียบวิธีสถิติทางการศึกษา 2. หลักสูตรและการสอน 3. จิตวิทยาทั่วไป หรือจิตวิทยาสำหรับครู

1.2 เนื้อหารายวิชา

ความสำคัญและความสัมพันธ์ระหว่างวัตถุประสงค์การศึกษา ประสพการณ์การเรียนรู้กับกระบวนการวัดและการประเมินผล หลักการวัดและการประเมินผลการเรียนการสอนตามทฤษฎีดั้งเดิม และทฤษฎีการตอบสนองข้อสอบ ประเภท ลักษณะ และการสร้างเครื่องมือการวัดและการประเมินผลทางการศึกษา เทคนิคการวิเคราะห์แบบสอบ การสร้างธนาคารข้อสอบ การให้อันดับคะแนน คะแนนมาตรฐาน ปกติวิสัยและการแปลงความหมายคะแนน หลักและวิธีการประเมินผลภายในและภายนอก เน้นระเบียบ แนวปฏิบัติ และการประยุกต์ในการวัดและการประเมินผลทางการศึกษา

1.3 วัตถุประสงค์เชิงพฤติกรรม

เมื่อนิสิตได้ศึกษาครบทุกเนื้อหาแล้ว จะต้องสามารถ

- อธิบายหลักการวัดและการประเมินผลการเรียนการสอนได้ถูกต้อง
- วิเคราะห์ข้อสอบและแบบสอบได้
- สร้างและพัฒนาแบบสอบผลสัมฤทธิ์ที่เชื่อถือได้
- ตรวจให้คะแนนและอันดับคะแนนได้
- ประยุกต์หลักการประเมินผลมาใช้ในการเรียนการสอนได้

1.4 วิธีจัดการเรียนการสอน

- บรรยาย
- อภิปราย
- ฝึกปฏิบัติแยกตามกลุ่ม
- ดูเทปการบรรยายพิเศษจากวิทยากร

1.5 สื่อการสอน

- แบบฝึก, ตัวอย่างเครื่องมือวัดผล, สื่ออิเล็กทรอนิกส์ใน www.edu.chula.ac.th/resched
- เทปการบรรยายพิเศษจากวิทยากร

1.6 การวัดผลการเรียน

- | | |
|--|-----|
| 1. สอบกลางภาค | 25% |
| 2. สอบปลายภาค | 25% |
| 3. รายงานการสร้างเครื่องมือประเมินผลการเรียนรู้ (งานกลุ่ม) | 30% |
| 4. การมีส่วนร่วมในชั้นเรียนและการเข้าชั้นเรียน | 10% |
| 5. การบ้าน | 10% |

1.7 เกณฑ์การประเมินผล

85% ขึ้นไป	ได้เกรด A
80% - 84%	ได้เกรด B+
75% - 79%	ได้เกรด B
70% - 74%	ได้เกรด C+
65% - 69%	ได้เกรด C
60% - 64%	ได้เกรด D+
55% - 59%	ได้เกรด D
ต่ำกว่า 55%	ได้เกรด F

ตอนที่ 2 การพิจารณาจำแนกข้อสอบ

จากวัตถุประสงค์ในการวิจัยครั้งนี้ นอกจากการพัฒนาคลังข้อสอบแล้ว สิ่งที่ต้องทำเป็นลำดับแรก คือ การจำแนกคุณลักษณะของข้อสอบวิชา 2702303 การวัดและการประเมินผลทางการศึกษาโดยพิจารณาจาก พฤติกรรมการเรียนรู้ ประเภทของเนื้อหาวิชา และคุณภาพของข้อสอบ รายข้อด้านความยากและอำนาจจำแนก ซึ่งมีรายละเอียดพิจารณาจำแนกคุณลักษณะข้อสอบแต่ละด้านดังนี้

2.1 พฤติกรรมการเรียนรู้

บลูม และคณะ (Bloom, B.S. et al., 1956 อ้างถึงใน ศิริชัย การญจนวาศี, 2544) ได้แบ่งระดับขั้นพัฒนาการทางสมองและสติปัญญา ดังนี้

1. ความรู้ความจำ (Knowledge)

ความรู้ความจำ หมายถึง ความสามารถทางสมองของผู้เรียนในการรับรู้ (ฟัง/อ่าน) รักษาความรู้ (จำ) และระลึก (ตอบ) ความรู้นั้นได้อย่างถูกต้องตามคำบรรยาย เอกสาร หรือตำรา สามารถจำแนกเป็นพฤติกรรมย่อยได้ ดังนี้

1.1 ความรู้ในเนื้อเรื่องเฉพาะ (Knowledge of specifics) ได้แก่

- 1.1.1 คำศัพท์ (Terminology) จำนิยาม สัญลักษณ์ เครื่องหมาย
- 1.1.2 ข้อเท็จจริง (Specific facts) จำข้อความจริง สูตร กฎ ทฤษฎี

1.2 ความรู้ในวิธีดำเนินการ (Knowledge of procedures) ได้แก่

- 1.2.1 ระเบียบแบบแผน (Convention) จำหลักประเพณี ระเบียบ ข้อตกลง
- 1.2.2 แนวโน้มและลำดับขั้น (Trends and sequences) จำขั้นตอนการปฏิบัติเปรียบเทียบความน่าจะเป็นของลำดับขั้นตอน

- 1.2.3 การจัดประเภทและระบบจำแนก (Classification and categories)
 - จำการจัดหมวดหมู่ ประเภท ชนิดของสิ่งต่าง ๆ
- 1.2.4 เกณฑ์ (Criteria) จำคุณสมบัติที่ใช้ในการจำแนกหรือตัดสิน
- 1.2.5 วิธีการ (Methodology) จำการดำเนินงานให้บรรลุตามเป้าหมาย
- 1.3 ความรู้ในหลักสากลและนามธรรม (Knowledge of universal and abstractions)
 - 1.3.1 หลักการและนัยทั่วไป (Principles and generalization) จำหลักการสำคัญและการสรุปอ้างอิง
 - 1.3.2 ทฤษฎีและโครงสร้าง (Theories and structure) จำทฤษฎีและโครงสร้างของคุณลักษณะตามหลักการหรือทฤษฎี
2. ความเข้าใจ (Comprehension)

ความเข้าใจ หมายถึง ความสามารถทางสมองของผู้เรียนในการเรียนรู้ จำ และสื่อสาร (บอก/เล่า/บรรยาย) ความรู้นั้นออกมาได้อย่างถูกต้อง สามารถจำแนกเป็นพฤติกรรมย่อยได้ ดังนี้

 - 2.1 การแปลความหมาย (Translation) บอกความหมายตามนัยของคำ/กิจกรรม
 - 2.2 การตีความ (Interpretation) นำผลการแปลความมาเปรียบเทียบเป็นข้อยุติ
 - 2.3 การขยายความ (Extrapolation) เปรียบเทียบความหมายของคำ/กิจกรรมที่กว้างไกลออกไปจากเดิม
3. การประยุกต์ (Application)

การประยุกต์ หมายถึง ความสามารถในการนำความรู้ ความเข้าใจที่มีอยู่ไปใช้ในการแก้ปัญหาของสิ่งนั้นในสถานการณ์ใหม่
4. การวิเคราะห์ (Analysis)

การวิเคราะห์ หมายถึง ความสามารถในการแยกแยะองค์รวมของสิ่งต่าง ๆ ออกเป็นส่วนประกอบย่อยๆ ได้ทำให้เห็นโครงสร้างของสิ่งนั้น สามารถจำแนกเป็นพฤติกรรมย่อยได้ ดังนี้

 - 4.1 การวิเคราะห์แบบแยกส่วนประกอบ (Analysis of elements) แยกแยะคุณลักษณะขององค์รวมเป็นส่วนประกอบย่อย
 - 4.2 การวิเคราะห์แบบเชื่อมโยงความสัมพันธ์ของส่วนประกอบ (Analysis of relationship) แยกแยะองค์รวมเป็นส่วนประกอบย่อยที่สัมพันธ์กัน
 - 4.3 การวิเคราะห์แบบเชื่อมโยงโครงสร้างของหลักการ (Analysis of organizational principles) แยกแยะองค์รวมเป็นโครงสร้างของส่วนประกอบที่สัมพันธ์กัน
5. การสังเคราะห์ (Synthesis)

การสังเคราะห์ หมายถึง ความสามารถในการผสมผสานส่วนประกอบย่อยเข้าด้วยกันเป็นองค์รวมใหม่ที่กลมกลืนอย่างมีความหมาย สามารถจำแนกเป็นพฤติกรรมย่อยได้ ดังนี้

- 5.1 การสังเคราะห์ข้อความ (Production of a unique communications) รวบรวมข้อความความเป็นข้อสรุปความสำคัญ
- 5.2 การสังเคราะห์แผนงาน (Production of a plan or operations) รวมส่วนประกอบย่อยเข้าด้วยกันเป็นแผนการดำเนินงานให้บรรลุผลตามเป้าหมาย
- 5.3 การสังเคราะห์แนวคิด (Derivation of abstract relations) ผสมผสานความรู้จากหลายแหล่ง แล้วสรุปเป็นแนวความคิด/องค์ความรู้อย่างเป็นระบบแบบแผน

6. การประเมิน (Evaluation)

การประเมิน หมายถึง ความสามารถในการตีค่า หรือตัดสินคุณค่าของสิ่งต่างๆ ตามเกณฑ์หรือมาตรฐานที่กำหนดไว้ สามารถจำแนกเป็นพฤติกรรมย่อยได้ ดังนี้

- 6.1 การประเมินค่าโดยใช้เกณฑ์ภายใน (Judgements in terms of internal criterion) ตัดสินคุณค่าตามเกณฑ์ภายในที่กำหนดไว้ในเรื่องนั้น
- 6.2 การประเมินค่าโดยใช้เกณฑ์ภายนอก (Judgements in terms of external criterion) ตัดสินคุณค่าตามเกณฑ์ภายนอก ที่กำหนดได้อย่างเป็นมาตรฐาน

2.2 ประเภทของเนื้อหาวิชา

ขอบเขตเนื้อหาวิชา 2702303 การวัดและการประเมินผลทางการศึกษา ที่ใช้เป็นกรอบในการออกข้อสอบมีดังนี้

1. หลักการวัดและการประเมินผล/การวัดและการประเมินผลตามพระราชบัญญัติการศึกษา
 - หลักการวัดและการประเมินผล
 - องค์ประกอบทางการศึกษา
 - ความหมายของ "การวัด" และ "การประเมิน"
 - ประเภทของการวัดและการประเมินผล
 - การวัดประเมินผลตามพระราชบัญญัติการศึกษา
 - ข้อกำหนดการวัดและการประเมินผลการเรียนตามหลักสูตรการศึกษาขั้นพื้นฐาน
 - หลักการวัดและการประเมินผลผลการเรียนตามหลักสูตรการศึกษาขั้นพื้นฐาน
 - แนวปฏิบัติการวัดและการประเมินผลการเรียนรู้ตามหลักสูตรการศึกษาขั้นพื้นฐาน
2. การประเมินผลการเรียนรู้
 - พฤติกรรมการเรียนรู้ : พุทธิพิสัย, จิตพิสัย, ทักษะพิสัย
 - วิธีการวัดผลการเรียนรู้

- จุดมุ่งหมายเชิงพฤติกรรม
- 3. เครื่องมือวัดผล
 - ความหมายของแบบสอบ
 - ประเภทของแบบสอบ
 - เครื่องมือการประเมินตามสภาพจริง (Authentic assessment) เช่น ทักษะการปฏิบัติ และ พอร์ตโฟลิโอ (Performance and portfolio assessment)
 - รูปแบบของข้อสอบ
- 4. การสร้างเครื่องมือวัดและประเมินผลการเรียนรู้
 - ขั้นตอนการสร้างข้อสอบวัดผลสัมฤทธิ์
 - ตารางวิเคราะห์หลักสูตร
 - การเขียนข้อสอบวัดพฤติกรรมระดับต่างๆ
 - ลักษณะเฉพาะของแบบสอบ
- 5. การตรวจสอบคุณภาพของเครื่องมือรายข้อ, ทฤษฎีดั้งเดิม
 - ความสำคัญของการวิเคราะห์ข้อสอบ
 - ความหมายของความยากและอำนาจจำแนก
 - การวิเคราะห์ข้อสอบแบบอิงกลุ่ม
 - การวิเคราะห์ข้อสอบแบบอิงเกณฑ์
 - ข้อตกลงเบื้องต้นและหลักการของทฤษฎีการทดสอบแบบดั้งเดิม
- 6. การตรวจสอบคุณภาพของเครื่องมือทั้งฉบับ
 - ความหมายของความตรงและความเที่ยง
 - ความตรง และวิธีตรวจสอบความตรง
 - องค์ประกอบที่มีผลต่อความตรง
 - ความเที่ยง และวิธีตรวจสอบความเที่ยง
 - ความคลาดเคลื่อนมาตรฐานของการวัด
 - องค์ประกอบที่มีผลต่อความเที่ยง
- 7. ทฤษฎี IRT และการสร้างธนาคารข้อสอบ
 - หลักและข้อตกลงเบื้องต้นของทฤษฎี IRT
 - การสร้างธนาคารข้อสอบ
- 8. การประเมินจิตพิสัย (Affective assessment)
 - การสร้างเครื่องมือวัดจิตพิสัย
 - ประเภทของเครื่องมือ : แบบสอบถาม แบบสัมภาษณ์

- การสร้างมาตรวัด Thurstone, Likert และ Osgood
- 9. การประเมินทักษะการปฏิบัติ (Performance evaluation)
 - การสร้างเครื่องมือประเมินทักษะการปฏิบัติ
 - วิธีการประเมิน : การสังเกต, การประเมินผลงาน
 - เกณฑ์การให้คะแนน Rubrics
- 10. การตัดสินผลการเรียนรู้
 - ความหมายของการให้ระดับคะแนน
 - องค์ประกอบที่ใช้ในการตัดเกรด
 - ลักษณะของการตัดเกรด
 - ประเภทของการตัดเกรด: อิงกลุ่ม และอิงเกณฑ์
 - การให้คะแนนแบบอิงพัฒนาการ
 - ความหมายของคะแนน
 - การเปรียบเทียบระดับคะแนน
- 11. การประเมินคุณภาพภายในและภายนอก
 - การประเมินคุณภาพภายใน
 - การประเมินคุณภาพภายนอก
 - มาตรฐานการศึกษา และตัวบ่งชี้

2.3 คุณภาพของข้อสอบรายข้อ; ความยากและอำนาจจำแนก

การตรวจสอบคุณภาพของข้อสอบรายข้อตามแนวคิดทฤษฎีการทดสอบแบบดั้งเดิม โดยทั่วไปจะพิจารณาเกี่ยวกับ ค่าความยาก และอำนาจจำแนก เป็นตัวบ่งชี้ที่สำคัญในการพิจารณาตัดสินคุณภาพข้อสอบแต่ละข้อ โดยมีรายละเอียดของสูตรในการคำนวณ และเกณฑ์การแปลความหมายคุณภาพข้อสอบแต่ละด้านดังนี้

ความหมายของความยาก อำนาจจำแนก และประสิทธิภาพตัวลวง

ความยาก (P: Difficulty indexes) หมายถึง สัดส่วนของจำนวนคนที่ตอบข้อสอบข้อนั้นได้ถูกต้องต่อจำนวนผู้เข้าสอบทั้งหมด ดังนั้นหากข้อสอบที่ผู้สอบตอบข้อสอบได้เป็นจำนวนมาก ก็จะได้ถือว่าเป็นข้อที่ง่าย และข้อสอบที่ผู้สอบตอบได้จำนวนน้อย ก็ถือว่าเป็นข้อสอบที่มีความยาก

อำนาจจำแนก (r หรือ D: Discrimination indexes) หมายถึง ความสามารถของข้อสอบในการจำแนก/แยก ความแตกต่างระหว่างผู้สอบที่มีผลสัมฤทธิ์ทางการเรียนต่างกัน

ประสิทธิภาพตัวลวง หมายถึง ความสามารถของตัวลวงในการจูงใจให้มีผู้เลือกตอบ และกลุ่มที่มีผลสัมฤทธิ์ต่ำเลือกตอบมากกว่ากลุ่มที่มีผลสัมฤทธิ์สูง

ความยากและอำนาจจำแนก มีรายละเอียดของสูตรการคำนวณ และเกณฑ์การแปลความหมาย ดังนี้

ความยากและอำนาจจำแนก มีรายละเอียดของสูตรการคำนวณ และเกณฑ์การแปลความหมาย ดังนี้ (ศิริชัย กาญจนวาสี, 2544; ดวงกมล ไตรวิจิตรคุณ, 2541)

สูตร

$$p = \frac{R_H + R_L}{N_H + N_L}$$

$$r = \frac{R_H - R_L}{N_H \text{ or } N_L}$$

เมื่อ	P	ค่าความยาก
	r	ค่าอำนาจจำแนก
	R _H	จำนวนคนที่ตอบถูกในกลุ่มสูง
	R _L	จำนวนคนที่ตอบถูกในกลุ่มต่ำ
	N _H	จำนวนคนทั้งหมดในกลุ่มสูง
	N _L	จำนวนคนทั้งหมดในกลุ่มต่ำ

เกณฑ์การพิจารณาคุณภาพข้อสอบ (ศิริชัย กาญจนวาสี, 2544; ดวงกมล ไตรวิจิตรคุณ, 2541)

ความยาก (p)	ความหมาย
0.00 – 0.19	ข้อสอบยาก ควรตัดทิ้งหรือ ปรับปรุง
0.20 – 0.39	ข้อสอบค่อนข้างยาก สามารถเก็บไว้ใช้ได้ต่อไป
0.40 – 0.60	ข้อสอบยากพอเหมาะ สามารถเก็บไว้ใช้ได้ต่อไป
0.61 – 0.80	ข้อสอบค่อนข้างง่าย สามารถเก็บไว้ใช้ได้ต่อไป
0.81 – 1.00	ข้อสอบง่าย ควรตัดทิ้ง หรือ ปรับปรุง
อำนาจจำแนก (r)	ความหมาย
ต่ำกว่า 0	ข้อสอบจำแนกไม่ได้ ควรตัดทิ้ง หรือ ปรับปรุง
0.00 – 0.19	ข้อสอบจำแนกได้ต่ำ ควรตัดทิ้ง หรือ ปรับปรุง
0.20 – 0.29	ข้อสอบจำแนกได้พอใช้ สามารถเก็บไว้ใช้ได้ต่อไป
0.30 – 0.39	ข้อสอบจำแนกได้ดี สามารถเก็บไว้ใช้ได้ต่อไป
0.40 ขึ้นไป	ข้อสอบจำแนกได้ดีมาก สามารถเก็บไว้ใช้ได้ต่อไป

ดังนั้นข้อสอบที่ดีและสามารถเก็บไว้ใช้ต่อไป จึงควรมีค่าความยาก อยู่ระหว่าง 0.20 – 0.80 และค่าอำนาจจำแนก ตั้งแต่ 0.20 ขึ้นไป

เกณฑ์การพิจารณาตัวลวง

ความยาก (p)	ความหมาย
เท่ากับ 0	ใช้ไม่ได้ เพราะไม่มีคนเลือก
มากกว่า 0	ใช้ได้ เพราะมีคนเลือก
อำนาจจำแนก (r)	ความหมาย
น้อยกว่า 0	จำแนกได้ เพราะลวงให้คนกลุ่มต่ำตอบมากกว่ากลุ่มสูง
เท่ากับ 0	จำแนกไม่ได้ เพราะกลุ่มสูงตอบเท่ากับกลุ่มต่ำ
มากกว่า 0	จำแนกไม่ได้ เพราะลวงให้กลุ่มสูงตอบมากกว่ากลุ่มต่ำ

ดังนั้นตัวลวงที่มีประสิทธิภาพ จะต้องมีความยาก มากกว่าศูนย์ และค่าอำนาจจำแนก น้อยกว่าศูนย์หรือติดลบ

หลังจากการวิเคราะห์คุณภาพข้อสอบรายข้อ ด้วยการคำนวณค่าความยาก อำนาจจำแนก ผู้สอนสามารถคัดเลือกข้อสอบที่มีคุณภาพคือมีค่าความยากอยู่ในช่วง 0.2-0.8 และค่าอำนาจจำแนก 0.2 ขึ้นไป ซึ่งสามารถบันทึกคุณภาพข้อสอบไว้ใน บัตรข้อสอบ (Card items) หรือในคอมพิวเตอร์ เพื่อรวบรวมข้อสอบที่มีคุณภาพ จัดเรียงตาม ปีการศึกษา เนื้อหาวิชา หรือ วัตถุประสงค์ เพื่อจัดเก็บใน คลังข้อสอบ ที่มีขีด ความปลอดภัย เพื่อนำไปใช้ต่อไป

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตอนที่ 3 คลังข้อสอบ

3.1 มโนทัศน์เกี่ยวกับคลังข้อสอบ

คำว่า คลังข้อสอบ (Item bank) เริ่มใช้ในประเทศอังกฤษในช่วงกลางปี ค.ศ. 1960 ซึ่งเป็นการจำแนกข้อสอบออกเป็นกลุ่มๆ ตามเกณฑ์ที่กำหนด อย่างเป็นระบบเช่นเดียวกับการจัดหนังสือในห้องสมุด เพื่อความสะดวกรวดเร็วในการคัดเลือกข้อสอบมาใช้ในการสร้างแบบวัด คำภาษาอังกฤษที่นำมาบรรยายถึงคลังข้อสอบมีหลายคำ เช่น Question banks, Item pool, Item collection, Item reservoirs, Test item libraries ซึ่งใช้ในความหมายเดียวกัน (Coppin, 1976; ภาวิณี ศรีสุขวัฒนานันท์, 2543)

Millan และ Arter (1984) ให้ความหมายของคลังข้อสอบว่า เป็นแหล่งจัดเก็บข้อสอบที่มีความเกี่ยวข้องสัมพันธ์กันจำนวนมาก เพื่อสามารถนำเอาไปใช้ได้สะดวก

ภาวิณี ศรีสุขวัฒนานันท์ (2543) กล่าวว่า คลังข้อสอบ หมายถึง การจัดเก็บข้อสอบจำนวนมากอย่างเหมาะสมเป็นระบบ และมีกระบวนการเรียกคืนกลับมาใช้ให้ปรับปรุงใหม่

อุทัย บุญประเสริฐ (2540) คลังข้อสอบ หรือ ธนาคารข้อทดสอบ หมายถึง ที่รวมของข้อสอบ ซึ่งมีไว้เพื่อประโยชน์ในการใช้ข้อทดสอบเหล่านั้น ตามวัตถุประสงค์ของการทดสอบในโอกาสต่อไป

กรมวิชาการ (2534) ได้ให้ความหมาย คลังข้อสอบของโรงเรียนว่า เป็นแหล่งรวบรวมข้อสอบรายข้อ ที่วัดได้ตรงตามจุดประสงค์การเรียนรู้แต่ละจุดประสงค์ ในแต่ละรายวิชา เพื่อให้ครูผู้สอนนำไปใช้ด้านการวัดผล และการสอนซ่อมเสริมในแต่ละจุดประสงค์ที่ต้องการและจำเป็นได้อย่างสะดวกรวดเร็ว

กล่าวโดยสรุป คลังข้อสอบ หมายถึง แหล่งรวบรวมและจัดเก็บข้อสอบที่มีคุณภาพ โดยมีการจัดจำแนกข้อสอบตามเนื้อหาวิชา พฤติกรรมการเรียนรู้ และคุณภาพข้อสอบด้านความยากและอำนาจจำแนก ซึ่งอาจมีการจัดเก็บในลักษณะเอกสารหรือการใช้คอมพิวเตอร์อย่างใดอย่างหนึ่ง

จุฬาลงกรณ์มหาวิทยาลัย

จำนวนข้อสอบเท่าใดจึงจะเหมาะสมที่จะจัดเก็บเป็นคลังข้อสอบ Millan และ Arter (1984) เสนอว่าหากใช้เพื่อวัดผลในแต่ละครั้ง จำนวนข้อสอบ 10 ข้อ แต่ถ้าใช้เพื่อวัดผลผู้เรียนเป็นเวลา 1 ชั่วโมง จำนวนข้อสอบควรเป็น 50 ข้อ และยังมีข้อสอบในคลังข้อสอบมากเท่าใดก็จะดีมากขึ้น เพื่อลดปัญหาการใช้ข้อสอบซ้ำที่ถี่มากเกินไป แต่อย่างไรก็ตามข้อสอบที่จัดเก็บต้องมีลักษณะ การเขียนข้อสอบหรือคำถามคำตอบที่ชัดเจน มีความตรงตามเนื้อหา และค่าสถิติแสดงคุณภาพข้อสอบที่เหมาะสม

โดยทั่วไปการจัดแบ่งประเภทข้อสอบเพื่อจัดเก็บในคลังข้อสอบจะใช้เนื้อหารายวิชาเป็นเกณฑ์ โดยแบ่งเนื้อหาออกเป็นหัวข้อย่อยๆ หรือการแบ่งข้อสอบตามวัตถุประสงค์ ซึ่งจัดแบ่งข้อสอบลดหลั่นตามระดับของเนื้อหาหรือวัตถุประสงค์รายวิชา โดยกำหนดเป็นรหัสประจำข้อสอบที่แสดงถึงเนื้อหาและวัตถุประสงค์ที่ต้องการวัดนั้น

3.2 ขั้นตอนการพัฒนาระบบคลังข้อสอบด้วยไมโครคอมพิวเตอร์

ศิริชัย กาญจนวาสี (2541) ได้แบ่งขั้นตอนของการพัฒนาระบบคลังข้อสอบด้วยไมโครคอมพิวเตอร์ไว้ 7 ขั้นตอนดังนี้

1. การวางแผน
 - 1.1 วิเคราะห์สภาพปัญหา
 - 1.2 กำหนดลักษณะคลังข้อสอบ
 - 1.3 กำหนดรหัสประจำข้อสอบ (Coding of item)
 - 1.3.1 ระดับหลักสูตร (Instructional level)
 - 1.3.2 ชื่อวิชา (Subject area)
 - 1.3.3 จุดมุ่งหมายการเรียนรู้การสอน (Objectives)
 - 1.3.4 ค่าสถิติของข้อสอบ (Item statistics)
2. ออกแบบโปรแกรมคอมพิวเตอร์
 - 2.1 เลือกภาษาโปรแกรมคอมพิวเตอร์
 - 2.2 เขียนโปรแกรมคอมพิวเตอร์
 - 2.3 ทดสอบโปรแกรม
 - 2.4 ประเมินระบบคลังข้อสอบ
 - 2.5 จัดทำคู่มือการใช้
3. การจัดเก็บข้อสอบเข้าคลัง
 - 3.1 การบันทึกข้อสอบเข้าคลัง
 - 3.2 การปรับปรุงแก้ไข

4. การคัดเลือกข้อสอบจากคลังมาใช้งาน
 - 4.1 กำหนดลักษณะแบบสอบที่ต้องการ
 - 4.2 คัดเลือกโดยการสุ่มข้อสอบ
 - 4.3 จัดเรียงข้อสอบ
5. การจัดพิมพ์ข้อสอบ
 - 5.1 กำหนดรูปแบบการจัดพิมพ์ข้อสอบ
 - 5.2 จัดพิมพ์ข้อสอบ
6. การให้สารสนเทศเกี่ยวกับคลังข้อสอบ
 - 6.1 การจำแนกลักษณะข้อสอบ
 - 6.2 จำนวนข้อสอบ ค่าสถิติของข้อสอบ
7. การรักษาความปลอดภัยของระบบคลังข้อสอบ
 - 7.1 การป้องกันการเข้าถึงคลังข้อสอบ
 - 7.1.1 แบ่งระดับผู้ใช้และการเข้าถึง
 - 7.1.2 กำหนดรหัสผ่าน และเปลี่ยนแปลงเป็นระยะ ๆ
 - 7.2 การป้องกันการทำลายข้อสอบ
 - 7.2.1 ซอฟต์แวร์ช่วยเตือน ตรวจ และทำลายสิ่งแปลกปลอม
 - 7.2.2 แผ่นต่อต้านไวรัส

Millan และ Arter (1984) เสนอแนวคำถามที่ผู้สร้างระบบการออกแบบคลังข้อสอบที่ต้องคำนึงถึง โดยการตั้งคำถามในประเด็นต่าง ๆ ดังต่อไปนี้

1. ตัวข้อสอบ

1.1 ความพอเพียงและการพัฒนาข้อสอบ

1.1.1 เราเป็นผู้ออกข้อสอบเองหรือจะใช้ข้อสอบของผู้อื่น หากออกข้อสอบเอง จะมีขั้นตอนการทำงานอย่างไร และถ้าใช้ข้อสอบของผู้อื่นจะขออนุญาตอย่างไร ข้อสอบที่มีผู้สร้างไว้แล้วจะมีเนื้อหาเหมาะสมกับการนำไปใช้หรือไม่

1.1.2 รูปแบบหรือประเภทข้อสอบมีลักษณะอย่างไร เป็นข้อสอบปลายเปิดหรือเป็นแบบกำหนดคำตอบให้ ถ้าเป็นข้อสอบแบบเลือกตอบจะมีกี่ตัวเลือก และข้อสอบแต่ละข้อจะมีการตรวจสอบคุณภาพความตรงหรือไม่ และมีค่าความยากอำนาจจำแนกเท่าใด

- 1.1.3 จำนวนข้อคำถามที่ต้องการจัดเก็บมีจำนวนเท่าใด โดยพิจารณาจำนวนข้อต่อเนื้อหาหรือวัตถุประสงค์
- 1.1.4 มีการปรับแก้ข้อสอบ หรือจัดพิมพ์ ก่อนนำเอาไปใช้อย่างไร ใครจะเป็นผู้ปรับปรุงแก้ไขข้อสอบ จะมีเกณฑ์อย่างไรในการจัดเก็บข้อสอบภายหลังที่ได้นำไปใช้แล้ว
- 1.2 การจำแนกข้อสอบ
- 1.2.1 กลุ่มเนื้อหาหรือคำสำคัญหรือทั้งสองอย่าง มีระดับชั้นของแต่ละหัวข้ออย่างไร และใครจะเป็นผู้จัดประเภทข้อสอบ
- 1.2.2 ข้อมูลเกี่ยวกับข้อสอบใดบ้างที่ต้องจัดเก็บในคลังข้อสอบ
- 1.2.3 มีการวัดเพื่อให้ได้สารสนเทศของข้อสอบที่ต้องการจัดเก็บอย่างไร
- 1.3 การจัดการ
- 1.3.1 มีการจัดการให้ข้อสอบมีความเป็นปัจจุบันอย่างไร ใครเป็นผู้ดำเนินการ มีขั้นตอนกระบวนการอย่างไร ข้อสอบมีค่าสถิติรายข้อเหมาะสมหรือไปข้อใดที่ควรลบออกไปจากคลังข้อสอบนั้น
- 1.3.2 มีจัดการกับข้อสอบที่มี รูปภาพ กราฟ ตัวอักษรลักษณะพิเศษ และอื่น ๆ ใด
- 1.3.3 มีจัดการกับข้อสอบที่มีลักษณะการถามตอบต่อเนื่อง หรือข้อสอบที่มีความเกี่ยวข้องกันอย่างไร
2. แบบสอบ
- 2.1 การรวบรวมข้อสอบ
- 2.1.1 มีการคัดเลือกข้อสอบแต่ละข้อ เพื่อจัดทำแบบสอบอย่างไร
- 2.1.2 หากข้อสอบที่มีการคัดเลือกด้วยระบบคอมพิวเตอร์ จะมีการสุ่มเลือกข้อสอบอย่างไร ข้อสอบแต่ละข้อใช้บ่อยแค่ไหน จะทำอย่างไรหากไม่มีข้อสอบที่ต้องการในเนื้อหานั้น
- 2.1.3 ค่าพารามิเตอร์ของข้อสอบหรือแบบสอบควรกำหนดเท่าไร กำหนดค่าความยากและความเที่ยงที่คาดไว้หรือไม่
- 2.1.4 มีกระบวนการในการจัดทำข้อสอบอย่างไร ใช้คำสั่งเลือกอย่างสุ่ม หรือการจัดให้ข้อสอบแตกต่างกันในแต่ละแบบสอบ
- 2.1.5 ใช้กระบวนการในการจัดพิมพ์แบบสอบอย่างไร
- 2.2 การบริหารการสอบ การให้คะแนน และการรายงานผล

- 2.2.1 มีการจัดการสอบด้วยคอมพิวเตอร์อย่างไร มีขั้นตอนการดำเนินการที่ขั้นตอนอย่างไรบ้าง
- 2.2.2 มีระบบการให้คะแนนหรือไม่ การให้คะแนนสำหรับข้อที่ถูกเพียงอย่างเดียวหรือให้คะแนนสำหรับค่าการเดา การให้น้ำหนักการตอบแต่ละข้ออย่างไร มีการประเมินผลการตอบโดยการเปรียบเทียบกับเฉลยหรือไม่
- 2.2.3 จะมีระบบรายงานผลการสอบอย่างไร เป็นการให้คะแนนเป็นรายบุคคลหรือเทียบคะแนนกับกลุ่มคนอื่นๆ รูปแบบของรายงานผลมีกี่รูปแบบ และจะรายงานผลไปให้ใครทราบบ้าง

2.3 การประเมิน

- 2.3.1 ค่าความเที่ยงและความตรงเท่าใดจึงจะจัดเก็บ ใครจะเป็นผู้จัดเก็บและจะจัดเก็บตัวข้อสอบอย่างไร
- 2.3.2 มีการจัดทำค่าปกติวิจัยหรือไม่ และมีการวัดแบบอิงกลุ่มด้วยวิธีใด

ระบบ

3.1 การจัดทำระบบและการพัฒนา

- 3.1.1 ใครจะเป็นผู้จัดทำระบบและพัฒนาคลังข้อสอบ
- 3.1.2 ระบบที่ทำขึ้นจะส่งผ่านไปยังผู้อื่นอย่างไร และมีระดับความมากน้อยในการเปิดเผยข้อมูลแค่ไหน

3.2 ลักษณะของซอฟต์แวร์และฮาร์ดแวร์

- 3.2.1 ระบบจะมีส่วนช่วยในการจะเก็บข้อสอบอย่างไร มีขั้นตอนคู่มือการใช้หรือไม่
- 3.2.2 เครื่องคอมพิวเตอร์ที่ใช้มีความจุหรือหน่วยความจำเท่าไร
- 3.2.3 ข้อสอบที่มีการจัดเก็บจำนวนมากๆ สามารถแบ่งแยกเพิ่มข้อมูลสำหรับผู้ใช้งานแต่ละคนอย่างไร
- 3.2.4 โครงสร้างของระบบคลังข้อสอบมีลักษณะอย่างไร ในเรื่องของฐานข้อมูล การใช้โปรแกรมในวัตถุประสงค์ต่างๆ
- 3.2.5 อุปกรณ์หรือเครื่องใช้ใดบ้างที่ต้องใช้
- 3.2.6 ผู้ใช้อุปกรณ์หรือคอมพิวเตอร์จะมีการถนอมรักษาการใช้งานของเครื่องมืออย่างไร
- 3.2.7 ใครจะเป็นผู้รับผิดชอบดูแลรักษาเครื่องมือ

3.3 การติดตามและมีกอบรม

- 3.3.1 ลักษณะของระบบจะมีการติดตามอย่างไร จำนวนของข้อสอบในแต่ละกลุ่มเนื้อหา กลุ่มผู้ใช้ จำนวนการปรับเปลี่ยนแก้ไข การกระจายความยาวและลักษณะของข้อสอบ
- 3.3.2 กระบวนการของระบบในการเปลี่ยนแปลงข้อมูลเป็นอย่างไร
- 3.4 การเข้าสู่ระบบและความปลอดภัย
 - 3.4.1 ใครจะสามารถใช้ข้อสอบหรือข้อมูลอื่นๆในคลังข้อสอบได้ ใครเป็นผู้ใช้แบบสอบ
 - 3.4.2 ผู้ใช้เข้าสู่ระบบคลังข้อสอบโดยตรงหรือผ่านต้องผ่านสื่อกลางใด
 - 3.4.3 ขั้นตอนการการรักษาความปลอดภัยของข้อมูลอะไรบ้าง
 - 3.4.4 เนื้อหาของข้อสอบจะจัดเก็บที่ไหน
 - 3.4.5 ใครสามารถเข้าไปดูรายงานผลการสอบได้บ้าง

4. การใช้งาน

4.1 โดยทั่วไป

- 4.1.1 ใครคือผู้ตัดสินใจใช้คลังข้อสอบ ผู้ใช้แบบสอบต้องการอะไรบ้าง
- 4.1.2 ใครเป็นผู้พัฒนาและใช้แบบสอบ ใครจะได้รับอนุญาตใช้ระบบคลังข้อสอบ
- 4.1.3 ระบบจะสามารถจัดการกับความคาดหวังและความต้องการในการใช้งาน
- 4.1.4 ผลลัพธ์ของระบบคลังข้อสอบที่ใช้บรรลุตามความวัตถุประสงค์หรือไม่
- 4.1.5 ทำอย่างไรให้ผู้ใช้อยอมรับและเพิ่มความน่าเชื่อถือของคลังข้อสอบให้เพิ่มขึ้น

4.2 ความก้าวหน้าในการสอน

- 4.2.1 คลังข้อสอบมีส่วนในการสอนหรือการตัดสินใจหรือไม่
- 4.2.2 หนังสือเรียน แนวหลักสูตร และเอกสารอื่นๆ ใดที่ใช้ในการสร้างข้อสอบ
- 4.2.3 ใครเป็นผู้จัดการตรวจสอบความตรง
- 4.2.3 มีข้อสอบที่ใช้สำหรับการฝึกฝนการสอบหรือไม่
- 4.2.4 มีข้อมูลที่จะช่วยให้ผู้ใช้งานสามารถวินิจฉัยเกี่ยวกับการตัดสินใจทางการศึกษาหรือไม่

4.3 การทดสอบแบบปรับเหมาะ

- 4.3.1 มีตารางการจัดการสอบอย่างไร
- 4.3.2 ข้อสอบที่คัดเลือกใช้ในแบบสอบจะมีประสิทธิภาพที่เป็นตัวแทนของเนื้อหาและไม่ก่อให้เกิดการใช้ซ้ำได้อย่างไร

- 4.3.3 เกณฑ์อะไรที่จะใช้ในการยุติการทำแบบสอบ
- 4.3.4 มีขั้นตอนการให้คะแนนอะไรบ้าง
- 4.4 การรับรองความสามารถ
 - 4.4.1 ข้อสอบในคลังที่ใช้วัดครอบคลุมทักษะที่สำคัญๆหรือไม่
 - 4.4.2 ความต้องการในการสอบผ่านกี่ครั้งที่อนุญาต และมีการติดตามกำกับอย่างไร
- 4.5 การประเมินหลักสูตรหรือโปรแกรม
 - 4.5.1 มีความเป็นไปได้ในการใช้ระบบการวัดผลสัมฤทธิ์ของผู้เรียนจำนวนมากในคุณลักษณะที่เฉพาะหลายๆอย่างหรือไม่
 - 4.5.2 ข้อสอบในคลังสามารถวัดครอบคลุมวัตถุประสงค์ของโปรแกรมหรือไม่ หรือวัดเนื้อหามากกว่าวัตถุประสงค์ที่กำหนด
 - 4.5.3 คลังข้อสอบให้ผลการวัดที่มีมาตรฐานเดียวกันเมื่อวัดต่างเวลากันหรือไม่
- 4.6 ความต้องการในการทดสอบและรายงานผลให้กับหน่วยงานภายนอก
 - 4.6.1 ระบบสามารถจัดการกับความต้องการของการประเมินโปรแกรมสำหรับการคัดเลือกนักเรียนที่จะให้ทุน การประเมินความต้องการทางการศึกษาหรือไม่
 - 4.6.2 ระบบสามารถจัดการทดสอบและรายงานที่ต้องการหรือไม่
- 5. ค่าใช้จ่าย
 - 5.1 ราคาที่พอเป็นได้
 - 5.1.1 ราคาอะไรบ้างที่ใช้ในการสร้างและการสนับสนุนระบบคลังข้อสอบ เช่น พื้นที่ เครื่องมือ เวลา เครื่องอำนวยความสะดวกต่าง ๆ
 - 5.1.2 ราคานี้สามารถจ่ายได้หรือไม่
 - 5.2 การเปรียบเทียบราคา
 - 5.2.1 ราคาของระบบคลังข้อสอบเปรียบเทียบกับระบบอื่น ๆ ที่มีเป้าหมายเดียวกันเป็นอย่างไร
 - 5.2.2 มีค่าใช้จ่ายพิเศษเพิ่มขึ้นบ้างหรือไม่

3.3 การกำหนดข้อมูลเกี่ยวกับข้อสอบ

Millan และ Arter (1984) ได้สรุปการกำหนดข้อมูลเกี่ยวกับข้อสอบที่จะจัดเก็บในคลังข้อสอบควรประกอบด้วย รายละเอียดต่อไปนี้

1. ระบุจำนวนข้อสอบ
2. รูปแบบอักษรที่ใช้
3. รูปแบบการตอบข้อสอบที่ถูกต้อง
4. ลักษณะข้อสอบที่เพิ่มเติม เช่น รูปภาพ กราฟ สัญลักษณ์พิเศษ
5. การอ้างอิงข้อมูลหรือคำตอบจากข้อหนึ่งไปอีกข้อหนึ่ง
6. ประเภทของข้อสอบ รูปแบบการตอบ เช่น ข้อสอบความเรียง หรือข้อสอบแบบถูกผิด
7. การพิมพ์คำแนะนำ
8. การเชื่อมโยงข้อคำถาม
9. แหล่งของข้อสอบ
10. คำสำคัญหรือการจำแนกข้อสอบ
11. การระบุผู้ที่จะคัดเลือกหรือจัดพิมพ์ข้อสอบ
12. พฤติกรรมการเรียนรู้ในแต่ละระดับ เช่น ความรู้ ความจำ ความเข้าใจ
13. การกำหนด แสดงตัวอย่าง และบริบทของข้อสอบ
14. การตัดลินผล
15. ระดับการศึกษา การให้เกรด
16. หนังสือเรียน แนวหลักสูตร และเอกสารต่างๆ ที่เป็นเนื้อหาการวัด
17. ระบบรักษาความปลอดภัย
18. ข้อเสนอแนะจากผู้ใช้

3.4 ลักษณะข้อสอบที่วัด

ลักษณะค่าสถิติข้อสอบที่นำไปวัดผลแล้ว และจะนำไปจัดเก็บเข้าคลัง Millan และ Arter (1984) มีข้อเสนอแนะดังนี้

1. ค่าความยาก
2. ค่าอำนาจจำแนก
3. ค่าการตอบสนองของข้อสอบที่เหมาะสมตามรูปแบบ (model) การวัด
4. ข้อมูลความถี่ในการตอบของข้อสอบแต่ละข้อ
5. ค่าความลำเอียง
6. เวลาที่ทำข้อสอบเสร็จโดยเฉลี่ย

7. ประวัติการใช้ข้อสอบ เช่น จำนวนนักเรียนที่ทดสอบ เวลาที่จัดการสอบ ความถี่ในการใช้ รหัสของผู้ที่ทำแบบทดสอบ
8. ระดับของความสามารถในการอ่าน

3.5 ประโยชน์ของคลังข้อสอบ

ศิริชัย กาญจนวาสี (2541) สุพัฒน์ สุกมลสันต์ (2539) และ Choppin (1985) ได้สรุปประโยชน์ของคลังข้อสอบไว้ดังนี้

1. ทำให้ได้ข้อสอบที่ตรงกับจุดมุ่งหมายและเนื้อหาของวิชา และเลือกใช้ได้อย่างเหมาะสมกับสถานการณ์การสอบแต่ละคลัง
2. ทำให้ได้ข้อสอบและแบบสอบที่มีคุณภาพน่าเชื่อถือจำนวนมาก
3. สร้างข้อสอบได้อย่างรวดเร็วและถูกต้อง
4. ใช้ข้อสอบซ้ำได้หลายครั้ง ด้วยการเลือกใช้ข้อสอบที่มีระบบ
5. ประหยัดงบประมาณ และเวลา
6. ได้ข้อสอบที่มีลักษณะเป็นคู่ขนาน

ภาวิณี ศรีสุขวัฒนานันท์ (2543) ได้กล่าวถึงประโยชน์ของคลังข้อสอบต่อบุคคล 4 กลุ่ม ดังนี้

1. ผู้สอน โดยคลังข้อสอบทำหน้าที่จัดเก็บข้อสอบที่มีคุณภาพ สอดคล้องกับวัตถุประสงค์การเรียนรู้ สามารถสร้างข้อสอบได้หลายชุด เพื่อให้ผู้เรียนในช่วงเวลาที่ต่างกัน ให้ผลการสอบที่สามารถเปรียบเทียบกันได้ สามารถออกแบบข้อสอบที่มีความยากและอำนาจจำแนกเหมาะสมกับการทดสอบเป็นรายบุคคล

2. นักการศึกษาพัฒนาหลักสูตร โดยปริมาณข้อสอบที่คุณภาพ มีค่าพารามิเตอร์ประจำข้อในคลังข้อสอบ สามารถพัฒนาสารสนเทศเพื่อการพัฒนาหลักสูตร และวิธีการเรียนการสอน ผลการสอบไม่เพียงแต่ประเมินผลการสอนของผู้สอนแต่ละคน แต่สามารถเปรียบเทียบผลการสอนระหว่างผู้สอนหลายคน เป็นประโยชน์ในการประเมินรูปแบบการสอน เพื่อปรับปรุงกระบวนการเรียนการสอนให้เหมาะสมยิ่งขึ้น

3. ผู้เรียน บัณฑิตและสภาพการณ์จำนวนมากที่ส่งผลต่อการทำข้อสอบของผู้เรียนให้เกิดความเคลื่อนไหวของการวัด เช่น การเดาคำตอบ ความสะเพร่า การพัฒนาคลังข้อสอบให้เหมาะสมกับลักษณะของผู้สอบ สามารถประมาณค่าความสามารถของผู้สอบด้วยการการสร้างคลังข้อสอบด้วยทฤษฎีการตอบสนองข้อสอบ (Item response theory) จะช่วยลดความคลาดเคลื่อนที่เกิดขึ้นได้

4. นักวิจัยพัฒนาการวัดความสามารถของคน ด้วยความก้าวหน้าของทฤษฎีการทดสอบแนวใหม่ และเทคโนโลยีคอมพิวเตอร์ ทำให้ข้อสอบมีค่าพารามิเตอร์ประจำข้อสอบที่ไม่เปลี่ยนแปลงตามกลุ่มผู้สอบ การดำเนินการสอบด้วยคอมพิวเตอร์สามารถจัดกระทำได้อย่างเป็นระบบ สามารถจัดสอบได้บ่อยครั้งมากขึ้นตามความพร้อม และวิเคราะห์แปลผลความสามารถผู้สอบได้อย่างรวดเร็ว

3.6 โปรแกรม MicroCAT

โปรแกรม MicroCAT เป็นระบบทดสอบด้วยคอมพิวเตอร์ที่มีขีดความสามารถของการพัฒนากระบวนการทดสอบตั้งแต่การสร้าง การจัดเก็บข้อสอบเป็นคลังข้อสอบ และสามารถจัดชุดแบบสอบจากข้อสอบที่มีอยู่ในคลัง ตลอดจนการสอบและการประมวลผล โดยอาศัยคอมพิวเตอร์

ภาวิณี ศรีสุขวัฒนานันท์ (2543) ได้ศึกษาพบว่าโปรแกรม MicroCAT มีข้อจำกัดในการจัดเก็บบันทึกข้อสอบภาษาไทยเข้าในคลังข้อสอบ จึงได้เขียนโปรแกรมย่อยประกอบบางส่วนเพิ่มเติม และปรับขั้นตอนการจัดเตรียมข้อสอบก่อนนำเข้าสู่การจัดเก็บในคลังข้อสอบ ดังนี้

ลักษณะของระบบการทดสอบ MicroCAT

MicroCAT Testing system เป็นระบบที่พัฒนาขึ้นบนคอมพิวเตอร์ ประกอบด้วย 4 ระบบย่อย ได้แก่

ระบบที่ 1 ระบบการพัฒนาข้อสอบและแบบสอบ

1.1 ระบบพัฒนาข้อสอบ (Item development) MicroCAT Graphic item banker (GIB) เป็นส่วนของโปรแกรมใหญ่ที่ให้ผู้พัฒนาจัดเก็บข้อสอบเข้าคลังและตรวจแก้ไขข้อสอบที่จัดเก็บไว้แล้ว การทำงานในส่วนนี้ผู้ใช้สามารถอ่านคำแนะนำบนจอภาพได้ และสามารถเก็บข้อสอบได้ 999 ข้อ การจัดเก็บข้อสอบทำได้ 3 ฟอรัม คือ (1) ฟอรัมข้อความทั้งหมด ไม่แทรกรูปภาพใช้กับจอภาพ Monochrome (ใช้กับข้อสอบภาษาไทยและภาษาอังกฤษ) (2) ฟอรัมข้อสอบที่มีรูปภาพแทรก (ใช้ได้เฉพาะข้อสอบภาษาอังกฤษเท่านั้น) (3) ฟอรัมที่เป็นข้อสอบที่จัดพิมพ์ต้นฉบับด้วยวิธีการใดก็ได้แล้ว scan เข้าไปใน Graphics modes

1.2 ระบบพัฒนาแบบสอบ (Test development) โปรแกรมของการพัฒนาแบบสอบเขียนด้วยภาษา The MCATL author language สามารถเชื่อมต่อผู้ใช้ในการสร้างแบบสอบตามความต้องการที่หลากหลาย แต่แบบสอบที่นิยมสร้างมี 3 ลักษณะ คือ

(1) แบบสอบแบบดั้งเดิม (Conventional test) สร้างโดยกำหนดกลุ่มข้อสอบและจำนวนข้อในแบบสอบตายตัวเพื่อใช้การสอบแบบดั้งเดิม ผลการสอบใช้การนำจำนวนข้อที่ถูกต้องแบบสอบชนิดนี้มาต่อการจัดทำ เหมาะสำหรับการสอบในการจัดกลุ่ม หรือเปรียบเทียบ

ความสามารถระหว่างกลุ่มผู้สอบ แต่ไม่มีประสิทธิภาพพอที่จะบอกระดับความสามารถที่แท้จริงของบุคคล

(2) แบบปรับให้เหมาะกับความสามารถรายบุคคล (Adaptive testing) แบบสอบนี้มีประสิทธิภาพในการจัดกลุ่มคน และบอกระดับความสามารถที่แตกต่างระหว่างบุคคลได้อย่างกว้างๆ หลักการสำคัญของการสร้างแบบสอบชนิดนี้จะออกแบบให้แบบสอบมีระดับความยากง่ายที่เหมาะสมกับความสามารถของผู้สอบแต่ละบุคคล ซึ่งทำให้ได้สารสนเทศจากการสอบมากพอที่บ่งชี้ความสามารถของบุคคลได้ชัดเจน กล่าวคือ มีสารสนเทศในการอธิบายความสัมพันธ์ระหว่างผู้สอบที่มีความยากกับข้อสอบที่ง่าย ผู้สอบที่มีความสามารถสูงกับข้อสอบที่ยาก

(3) แบบสอบที่กำหนดจำนวนข้อและลักษณะค่าตัวแปรหรือสร้างแบบสอบที่ใช้แนวคิด Stradaptive โดยให้มีสารสนเทศจากการสอบสูงสุด (Maximum-information) หรือ Bayesian adaptive tests

แบบสอบที่สร้างขึ้นจะทำการแปลผล (Compile) เป็นภาษาเครื่องทั้งตัวคำชี้แจง และตัวข้อสอบทั้งฉบับก่อนนำไปใช้สอบ วิธีนี้เป็นการจัดชุดแบบสอบตามความต้องการ หรือวัตถุประสงค์ของการทดสอบแต่ละครั้งได้อย่างถูกต้อง แบบสอบที่สร้างขึ้นจากการประมวลข้อสอบที่คัดไว้และเติมคำสั่งไว้ ผู้ใช้สามารถจัดการสอบโดยใช้คอมพิวเตอร์หรือพิมพ์ลงในกระดาษแล้วทำการสอบแบบเดิม คือ เขียนตอบลงในกระดาษคำตอบย่อทำได้

เหตุผลของการแปลแบบสอบที่สร้างขึ้นเป็นภาษาเครื่อง มีดังนี้

- (1) เพื่อเป็นการตรวจสอบว่าข้อสอบที่ต้องการมีอยู่ในคลัง และตรวจสอบคำสั่งที่ใช้ด้วย
- (2) เพื่อความปลอดภัย เพราะไม่สามารถเปลี่ยนแปลงในระหว่างที่มีการสอบ
- (3) เพื่อการประหยัดเนื้อที่ เนื่องจากการแปล จะบีบอัดข้อมูลได้
- (4) เพื่อความรวดเร็วในการเสนอข้อสอบ ขณะที่ทำการสอบ

ระบบที่ 2 ระบบการสอบผ่านเครื่องคอมพิวเตอร์

ระบบการสอบเป็นระบบอัตโนมัติของการจัดสอบแล้วบันทึกคะแนนของผู้สอบลงในแฟ้มข้อมูล และสามารถรายงานผลสอบทันที วิธีการบริหารการสอบเป็นไปตามข้อกำหนดของแบบสอบแต่ละชนิด

ระบบที่ 3 ระบบการประเมิน (Assessment subsystem)

ระบบการประเมินประกอบด้วย โปรแกรมต่างๆ 6 ชุด

- (1) Data collection program ใช้เก็บข้อมูลที่เป็นผลการสอบจากจอโดยอัตโนมัติ มีไฟล์ Output เป็นนามสกุล .Dat

(2) The MicroCAT item and test analyses program (ITEMAN) ให้ผลวิเคราะห์แบบดั้งเดิมซึ่งจะช่วยพิจารณาในด้านความเที่ยงของแบบสอบจากผลการวิเคราะห์รายข้อและให้ค่าสถิติบรรยาย เช่น ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ความเบ้ ลักษณะการแจกแจง เป็นต้น

(3) Rash model item calibration program ให้ประมาณความสามารถในโมเดล IRT ของ Rash การวิเคราะห์ข้อสอบตามแนวทฤษฎีการตอบสนอง (IRT) แบบราล์ชโมเดลมีค่าพารามิเตอร์ประจำข้อสอบเพียงค่าเดียว คือ ค่าความยากประจำข้อสอบ (b) โปรแกรม MicroCAT ได้พัฒนาการวิเคราะห์ข้อสอบแบบราล์ชโมเดล ที่มีค่าพารามิเตอร์ประจำข้อที่คงที่มากกว่าการวิเคราะห์แบบโลจิสติกโมเดล

(4) The MicroCAT item parameter estimation program (ASCAL) ให้ค่าประมาณความสามารถของผู้สอบ และค่าพารามิเตอร์ของข้อสอบรายข้อ เช่น a b c เป็นต้น การประมาณค่าข้อสอบด้วยโมเดลโลจิสติก 2-3 พารามิเตอร์ ใช้แนวคิดของทฤษฎีการประมาณค่าความเป็นไปได้สูงสุด (Maximum likelihood) รวมกับการประมาณค่าแบบเบย์ (Bayesian) ค่าพารามิเตอร์ประจำข้อสอบที่ได้มี 3 ค่า คือ ค่าอำนาจจำแนก (a) ค่าความยาก (b) และค่าการเดา (c) และการประมาณค่าความสามารถของผู้สอบ (θ) อยู่ระหว่างค่า -3 ถึง 3

ระบบที่ 4 การสอบแบบดั้งเดิม

การพิมพ์แบบสอบที่สร้างจากระบบที่ 2 แล้วให้ผู้สอบตอบโดยใช้ปากกา/ดินสอ แล้วนำผลสอบเก็บบันทึกไว้ในข้อมูล เพื่อนำไปวิเคราะห์ผลการสอบต่อไป

การปรับเพื่อใช้ในการพัฒนาระบบการสอบในภาษาไทย

1. การจัดเก็บข้อสอบ

ปัญหาการจัดเก็บข้อสอบด้วยวิธีการจัดเก็บแบบตัวอักษรล้วน และวิธีการจัดเก็บแบบตัวอักษรและรูปภาพของ MicroCAT ไม่สามารถจัดเก็บภาษาไทย การแก้ปัญหาด้วยการใส่ Driver โปรแกรมภาษาไทยจะช่วยได้เฉพาะข้อความที่ไม่ใช่สูตรคณิตศาสตร์ แต่อย่างไรก็ตามพบว่าโปรแกรมไม่มีระบบการจัดหน้าเอกสาร ให้นำเสนอนหน้าจอที่มีคุณภาพน่าพอใจ

วิธีการที่เลือกสำหรับระบบพัฒนาการจัดเก็บข้อสอบแบบภาษาไทยจากการทดลอง พบว่าการจัดเก็บแบบภาษาไทยโดยใช้การ Scan สามารถเก็บข้อสอบที่เป็นทั้งภาพและตัวอักษรได้เป็นการ Scan ข้อสอบที่ได้จัดเตรียมด้วยกรรมวิธีต่างๆ ที่ให้ได้ต้นฉบับข้อสอบที่น่าพอใจที่สุดก่อนแล้ว Scan เข้าไปเป็นรูปภาพ จัดเก็บใน file ที่มีนามสกุล .Tfe ให้มีขนาดประมาณ 25K แล้วใช้คำสั่ง Add ข้อสอบใน Graphic mode ตามวิธีการจัดเก็บข้อสอบ

จุดอ่อนที่พบจากการใช้วิธีดังกล่าว คือ ต้องใช้พื้นที่เก็บข้อมูลค่อนข้างสูงเมื่อเทียบกับการจัดเก็บที่เป็นข้อความล้วน แต่เนื่องจากปัจจุบัน Hard disk มีราคาลดลงโดยที่มีความจุสูงขึ้น จึงไม่เป็นเรื่องน่าวิตก

2. การทำเมนู

เป็นการพัฒนากรอบเมนูโปรแกรมของ MicroCAT ให้เป็นภาษาไทย ถ้าเป็นการพัฒนา ระบบปฏิบัติการ Dos ต้องใช้โปรแกรม Driver ภาษาไทยเข้ามาช่วย แต่ถ้าเป็นระบบปฏิบัติการ windows 95 นั้น มี Driver ภาษาไทยมาให้ ในการแก้ไขเมนูให้เป็นภาษาไทยนั้นยังสามารถแก้ไขให้เป็นภาษาไทยได้ไม่สมบูรณ์เพราะบางส่วนไม่สามารถแก้ไขให้เป็นภาษาไทยได้ การแก้ไขเมนูให้เป็นภาษาไทยสามารถแก้ไขได้ที่ไฟล์ Menudata.Mic

3. การทำ Template แบบแผนการสร้างแบบสอบถามเป็นภาษาไทย

การพัฒนา Template ภาษาไทยนั้นเป็นการพัฒนายูสเซอร์อินเตอร์เฟซของโครงแบบสอบถาม ต่าง ๆ ให้เป็นภาษาไทยโดยทำการแก้ไขที่ไฟล์นามสกุล .Tpl เพื่อให้เกิดความสะดวกในการทำงาน

4. การปรับปรุง Report

แต่เดิมการรายงานผลการทดสอบเป็นการรายงานพร้อมทั้งข้อที่ตอบถูกต้องผิด แล้วสรุปเป็นเปอร์เซ็นต์ข้อที่ถูก และยังมี การแสดงหมายเลขข้อสอบในคลังข้อสอบด้วยผู้พัฒนาจึงทำการแก้ไขการรายงานผลการสอบให้เป็นการรายงานเฉพาะเปอร์เซ็นต์ที่ทำได้และการแสดงให้เป็นภาษาไทย และไม่แสดงหมายเลขประจำข้อ

3.7 งานวิจัยที่เกี่ยวข้อง

พลกร กรพิทักษ์ (2533) ได้ดำเนินการวิจัยเรื่อง การพัฒนาโปรแกรมคอมพิวเตอร์สำหรับการจัดเก็บข้อสอบโดยใช้ไมโครคอมพิวเตอร์ มีขั้นตอนในการดำเนินการประกอบด้วย การศึกษา ปัญหาของระบบการจัดเก็บข้อสอบที่ใช้อยู่ในปัจจุบัน วิเคราะห์ปัญหา ออกแบบโปรแกรมคอมพิวเตอร์ เลือกภาษาคอมพิวเตอร์และเขียนโปรแกรม ทดสอบโปรแกรม จัดทำเอกสารประกอบการใช้โปรแกรม และประเมินคุณภาพของโปรแกรม

ผลการวิจัยและพัฒนาทำให้ได้โปรแกรมคอมพิวเตอร์ระบบการจัดการ เอ็มเอส/พีซี-ดอส (MS/PC-DOS) Version 2 สำหรับการจัดเก็บ และเลือกสรรข้อสอบที่สามารถจัดเก็บและสะสมข้อสอบได้ตามวัตถุประสงค์ของการวิจัย โดยโปรแกรมดังกล่าว สามารถเลือกสรรข้อสอบที่จัดเก็บโดยอาศัยค่าพารามิเตอร์ของข้อกระทงมาเป็นเกณฑ์ในการคัดเลือก และยังสามารถจัดพิมพ์ข้อสอบที่เลือกสุ่มออกมาใช้เป็นแบบสอบถามได้อย่างเป็นระบบ จากการประเมินโปรแกรมสรุปได้ว่าโปรแกรมการจัดเก็บข้อสอบที่พัฒนาขึ้น มีความสามารถในการบันทึกข้อมูลของผู้ใช้โปรแกรมได้

โปรแกรมดังกล่าวยังมีความถูกต้องในการใช้งานตามต้องการและมีความเชื่อถือได้ในการใช้งาน รวมทั้งยังมีความทนทานต่อการผิดพลาดของผู้ใช้โปรแกรม

จากการประเมินผลของผู้ใช้โปรแกรม พบว่า ผู้ใช้โปรแกรมเห็นด้วยอย่างยิ่งว่า โปรแกรมที่พัฒนามีความง่ายต่อการใช้งานในขั้นการจัดเก็บและเพิ่มข้อสอบ และสามารถนำมาใช้แทนระบบการจัดเก็บข้อสอบแบบกระดาษได้เป็นอย่างดี ซึ่งมีประโยชน์ต่อการวัดผลทางการศึกษาโดยส่วนรวม ผู้ใช้โปรแกรมยังเห็นด้วยว่า คู่มือการใช้โปรแกรมมีความชัดเจนและสอดคล้องกับการใช้งานโปรแกรม โดยที่ผู้ใช้โปรแกรมไม่จำเป็นต้องมีความรู้พื้นฐานทางด้านคอมพิวเตอร์ พิมพ์ดีด หรือการศึกษา ก็สามารถที่ใช้โปรแกรมได้อย่างมีประสิทธิภาพ

ประคอง กรรณสูต (2536) ได้ศึกษา การพัฒนาคลังข้อสอบด้วยโปรแกรมประมวลผลคำ โดยใช้ไมโครคอมพิวเตอร์ ด้วยโปรแกรมประมวลผลคำราชวิถี Word PC Version 1.1 และไมโครคอมพิวเตอร์ตั้งแต่ 16 บิต ขึ้นไป ในการจัดเก็บข้อสอบวิชา 412103 สถิติการศึกษาชั้นนำ ที่ได้สร้างและปรับปรุงจนเป็นข้อสอบที่เป็นมาตรฐาน เก็บข้อสอบ 1 ข้อ ไว้ใน 1 แฟ้มข้อมูล โดยใช้ชื่อที่มีความหมายทั้งวัตถุประสงค์ในการวัดและเนื้อหาตามหมวดที่กำหนดไว้ ตัวอย่างประชากรเป็นแบบสอบ 4 ชุด ๆ ละ 12 ข้อ ต่างมีลักษณะเป็นคู่ขนาน ซึ่งได้มาโดยการสุ่มอย่างเจาะจงจากประชากรข้อสอบในคลังข้อสอบ และใช้ตารางวิเคราะห์เนื้อหาเป็นแนวทางในการเลือกข้อสอบ เพื่อให้แบบสอบมีความตรงตามเนื้อหา ข้อค้นพบและผลที่ได้จากการพัฒนาคลังข้อสอบ มีดังนี้

1) ได้คลังข้อสอบในแผ่นบันทึกข้อมูลวิชา 412103 สถิติการศึกษาชั้นนำ ข้อสอบแต่ละข้อในคลังข้อสอบมีชื่อที่มีความหมายทั้งวัตถุประสงค์ในการวัด และเนื้อหาตามหมวดที่กำหนดไว้ และถูกบันทึกไว้ 1 ข้อ ใน 1 แฟ้มข้อมูล คลังข้อสอบนี้เก็บรักษาง่าย เรียกมาใช้ได้ทันที เก็บและสร้างข้อสอบชนิดเดียวกันได้ทุกแบบ 2) ได้วิธีสร้างแบบทดสอบซึ่งผู้สอนจะทำได้เอง ดังนี้ 2.1 ผู้สอนสามารถเลือกข้อสอบได้ด้วยตนเอง 2.2 จุดเฉพาะชื่อแฟ้มข้อมูลของข้อสอบจากที่พิมพ์ไว้เป็นเอกสารตรวจสอบ 2.3 ใช้คำสั่ง ^KR เรียกชื่อแฟ้มข้อมูลจากข้อ 2.2 ขึ้นบนจอภาพ 2.4 บันทึกแบบทดสอบที่ปรับให้อยู่ในแบบที่ต้องการแล้วจากจอภาพลงในแผ่นบันทึกข้อมูลและเอกสารตรวจสอบและสั่งให้เครื่องพิมพ์ผ่านเครื่องพิมพ์ลงในกระดาษ และ 3) พนักงานพิมพ์ดีดก็สามารถผลิตแบบทดสอบตามคำสั่งตามคำสั่งของผู้สอนได้

วิบูลาศ เจริญชัย (2537) ได้พัฒนาโปรแกรมไมโครคอมพิวเตอร์จัดเก็บและวิเคราะห์ข้อสอบแบบเลือกตอบ โดยมีวิธีดำเนินการ 7 ขั้นตอน ได้แก่ 1) ศึกษาและวิเคราะห์ปัญหาการจัดเก็บและวิเคราะห์ข้อสอบ 2) ออกแบบโปรแกรม 3) เลือกภาษาคอมพิวเตอร์และเขียนโปรแกรม ด้วยภาษา FoxPro 4) ทดสอบโปรแกรม 5) การทำเอกสารประกอบโปรแกรม และ 7) การประเมินประสิทธิภาพโปรแกรม

กลุ่มตัวอย่างเป็นครูโรงเรียนมัธยมศึกษา จังหวัดนครราชสีมา จำนวน 42 คน ผลการวิจัยพบว่า โปรแกรมไมโครคอมพิวเตอร์จัดเก็บและวิเคราะห์ข้อสอบแบบเลือกตอบที่สามารถวิเคราะห์ข้อสอบแบบอิงกลุ่ม จัดเก็บข้อสอบที่มีคุณภาพ แก้ไข และคัดเลือกข้อสอบที่จัดเก็บได้ตามเกณฑ์การคัดเลือก รวมทั้งจัดพิมพ์ข้อสอบที่คัดเลือกมาใช้เป็นแบบสอบได้อย่างมีประสิทธิภาพ ผลการประเมินโปรแกรม พบว่า โปรแกรมมีประสิทธิภาพ เท่ากับ 91.25 ซึ่งสูงกว่าเกณฑ์ที่กำหนด และครูอาจารย์มีความความเห็นไปในทางที่ดีต่อโปรแกรมนี้ คู่มือการใช้โปรแกรมมีความชัดเจน และสอดคล้องกับการใช้งานของโปรแกรม ครูอาจารย์สามารถใช้โปรแกรมได้อย่างมีประสิทธิภาพ โปรแกรมมีประโยชน์ต่อการวัดผลทางการศึกษา สามารถใช้แทนการจัดเก็บข้อสอบแบบกระดาษได้เป็นอย่างดี

กอบกิจ สหัสรังษี (2539) ได้พัฒนาระบบการจัดการข้อสอบปรนัย สำหรับเครื่องไมโครคอมพิวเตอร์และสามารถใช้ในระบบเครือข่ายท้องถิ่น โดยระบบการจัดการข้อสอบปรนัยที่พัฒนาขึ้นประกอบด้วย 2 ส่วน คือ ส่วนระบบด้านครู และส่วนระบบด้านการจัดการด้านนักเรียน

ในส่วนแรกเป็นส่วนที่ทำหน้าที่จัดแฟ้มข้อมูลหลักต่างๆ ของระบบ เช่น การสร้างโครงสร้างแฟ้มข้อมูล ปรับปรุงข้อมูล สร้างชุดข้อสอบ ชนิดต่าง ๆ โดยสามารถสร้างข้อสอบได้ 4 ชนิด คือ ข้อสอบ ชนิดเลือกตอบ ข้อสอบชนิดถูกผิด ข้อสอบชนิดจับคู่ และข้อสอบชนิดเติมคำ ซึ่งข้อสอบทั้ง 4 ชนิด สามารถแสดงผลได้ทั้งตัวอักษรภาษาไทยและภาษาอังกฤษ ส่วนข้อสอบชนิดเลือกตอบและชนิดถูกผิดนั้นสามารถแสดงผลออกเป็นได้ทั้งภาพและเสียง โดยที่ข้อสอบทั้ง 4 ชนิดได้รับการออกแบบให้สามารถสุ่มเพื่อสลับทั้งข้อสอบและตัวเลือกได้โดยระบบจัดการอัตโนมัติ นอกจากนี้ยังช่วยจัดการเกี่ยวกับการวัดผลให้คะแนนและเกรดได้ รวมทั้งการวิเคราะห์ข้อสอบด้วย

ส่วนที่สอง เป็นส่วนที่ทำหน้าที่แสดงกระดาษข้อสอบ ให้นักเรียนทำการสอบหรือแบบสอบจากข้อสอบทั้ง 4 ชนิดที่ครูเป็นผู้กำหนด โดยการทำการสอบไม่ว่าจะเป็นข้อสอบชนิดใด ผู้สอบจะเห็นภาพการทำข้อสอบเหมือนกับทำจากกระดาษจริง เพียงแต่ใช้เมาส์กดและลากเท่านั้น สามารถเลือกทำข้อไหนก่อนก็ได้และสามารถย้อนกลับมากแก้ไขข้อใดก็ได้ สุดท้ายระบบยังแสดงเวลาที่เหลือให้เห็นและจะทำการตัดการทำงานเมื่อหมดเวลาอีกด้วย

การวิจัยของ กอบกิจ สหัสรังษี (2539) นี้มีประโยชน์ต่อระบบการเรียนการสอน ช่วยให้ผู้สอนใช้คอมพิวเตอร์ สร้างแบบสอบและข้อสอบ พร้อมทั้งทำการวัดผลการเรียนการสอนได้อย่างสะดวกรวดเร็วและมีประสิทธิภาพมากยิ่งขึ้น ขณะเดียวกันก็เป็นประโยชน์ต่อผู้สอนที่จะนำแบบสอบหรือข้อสอบนั้น ๆ มาปรับปรุงหรือเพิ่มเติมและนำกลับมาใช้ทดสอบใหม่ได้หลาย ๆ ครั้ง

กฤษญาพันธ์ พงษ์บริบูรณ์ (2543) ได้พัฒนาโปรแกรมคอมพิวเตอร์ เพื่อการวิเคราะห์ข้อสอบและการจัดเก็บ สำหรับแบบทดสอบแบบอิงเกณฑ์ พร้อมทั้งประเมินประสิทธิภาพของโปรแกรมคอมพิวเตอร์ที่พัฒนาขึ้น กลุ่มเป้าหมายที่ใช้สำหรับประเมินคุณภาพของโปรแกรมคอมพิวเตอร์เป็นครูโรงเรียนมัธยมศึกษา จังหวัดขอนแก่น การดำเนินการวิจัยมี 2 ระยะ คือ ระยะที่

1 เป็นการพัฒนาโปรแกรมสำหรับวิเคราะห์ข้อสอบ และการจัดเก็บข้อสอบของแบบสอบแบบอิงเกณฑ์ ซึ่งมีขั้นตอนดังต่อไปนี้ 1) กำหนดปัญหาและความต้องการ 2) การวิเคราะห์ปัญหาและความต้องการ 3) การออกแบบโปรแกรม 4) การเลือกใช้ภาษา FoxPro และเขียนโปรแกรม 5) การทดสอบโปรแกรม 6) การจัดทำเอกสารประกอบการใช้โปรแกรม และระยะที่ 2 7) เป็นการประเมินประสิทธิภาพของโปรแกรมคอมพิวเตอร์ สำหรับการวิเคราะห์ข้อสอบและจัดเก็บข้อสอบของแบบสอบแบบอิงเกณฑ์

ผลการวิจัยได้โปรแกรมคอมพิวเตอร์ สำหรับวิเคราะห์ข้อสอบแบบอิงเกณฑ์และจัดเก็บข้อสอบของแบบทดสอบแบบอิงเกณฑ์หรือโปรแกรม ITB 1.00 สำหรับข้อสอบแบบปรนัยชนิดเลือกตอบ ซึ่งไม่เกิน 5 ตัวเลือก ที่ใช้วิธีการให้คะแนนแบบ 1-0 และมีจำนวนข้อสอบในแบบสอบไม่เกิน 120 ข้อ สำหรับโปรแกรม ITB 1.00 นี้ ผู้วิจัยใช้โปรแกรมสำเร็จรูป FoxPro R 2.6 ในการเขียนโปรแกรม ขั้นตอนการทำงานของโปรแกรม คือ การเตรียมฐานข้อมูล การประมวลผล และการรายงานผลทางจอภาพและเครื่องพิมพ์ ผลการวิเคราะห์ข้อสอบแบ่งออกเป็น 3 ส่วน ดังนี้ 1) ค่าสถิติเบื้องต้นของแบบทดสอบ คะแนนรวม คะแนนในแต่ละจุดประสงค์ และการตัดสินการผ่าน/ไม่ผ่านจุดประสงค์ จำนวนข้อสอบ จำนวนนักเรียน ค่าเฉลี่ย ค่าเบี่ยงเบนมาตรฐาน 2) คุณภาพแบบสอบ ค่าความเที่ยงของแบบทดสอบ ค่าความยาก (P) และค่าอำนาจจำแนก (ดัชนี B) ของข้อสอบรายข้อ (3) ข้อที่ผ่านการคัดเลือกว่าเป็นข้อสอบที่มีคุณภาพและข้อสอบที่ใช้ไม่ได้

การทดสอบโปรแกรมด้านความตรง ความเชื่อมั่นและความปลอดภัย โดยผู้เชี่ยวชาญพบว่า โปรแกรมที่พัฒนาขึ้นมีความตรงเชิงโครงสร้าง โปรแกรมมีความเชื่อมั่นด้านผลการวิเคราะห์ของโปรแกรมอย่างสม่ำเสมอและโปรแกรมให้ความปลอดภัยด้านข้อมูล และระบบโปรแกรมได้ การประเมินประสิทธิภาพโปรแกรม โดยผู้ใช้โปรแกรม พบว่า คู่มือโปรแกรมฐานข้อมูล การประมวลผล การรายงานผลและวิธีใช้โปรแกรม ในส่วนการใช้ภาษาที่ปรากฏบนจอภาพอย่างชัดเจน คำแนะนำสามารถปฏิบัติตามได้ กระบวนการใช้งานง่าย และรูปแบบการนำเสนอเข้าใจง่าย

วีรยุทธ์ ธานี (2546) ได้พัฒนาโปรแกรมคลังข้อสอบบนระบบปฏิบัติการ Windows เพื่อให้สามารถจัดเก็บข้อสอบเลือกตอบที่เป็นข้อความ รูปภาพ และสัญลักษณ์ รวมทั้งเป็นระบบที่สามารถบันทึก คัดเลือก แสดงผล และจัดพิมพ์ได้ พร้อมทั้งสร้างคู่มือในการใช้โปรแกรมคลังข้อสอบ และประเมินผลระบบการทำงานของโปรแกรมคลังข้อสอบ โดยมีการดำเนินการเป็น 3 ขั้นตอน คือ ขั้นตอนที่ 1 ศึกษาสภาพปัญหาและความต้องการระบบคลังข้อสอบ ขั้นตอนที่ 2 พัฒนาระบบคลังข้อสอบ และขั้นตอนที่ 3 ประเมินผลระบบคลังข้อสอบ

ผลการวิจัยในแต่ละขั้นตอนมีดังนี้ ขั้นตอนที่ 1 ศึกษาสภาพปัญหาและความต้องการระบบคลังข้อสอบด้วยการสัมภาษณ์ครูอาจารย์ในสังกัดสำนักงานประถมศึกษาจังหวัดอุบลราชธานี จำนวน 20 คน จาก 10 โรงเรียน พบว่า ในด้านความต้องการระบบคลังข้อสอบ

ส่วนใหญ่มีความต้องการระบบคลังข้อสอบที่เก็บด้วยระบบคอมพิวเตอร์และเป็นโปรแกรมที่ใช้กับระบบปฏิบัติการ Windows ส่วนการกำหนดลักษณะข้อสอบที่ต้องการจะยึดจุดประสงค์การเรียนรู้และเนื้อหา ในด้านคุณลักษณะของข้อสอบที่ต้องการเก็บส่วนใหญ่ต้องการเก็บข้อสอบที่เป็นตัวอักษรและรูปภาพ มีจำนวนตัวเลือก 4 ตัวเลือก และหัวข้อตัวเลือก มักเป็น ก ข ค ง ยกเว้นวิชาภาษาอังกฤษที่ต้องการหัวข้อตัวเลือกเป็น a b c d ในด้านของรายละเอียดของข้อสอบที่ต้องการเก็บ ได้แก่ วิชา จุดประสงค์ เนื้อหา ความยากง่าย อำนาจจำแนก ค่าการเดา พฤติกรรมที่สอบ และเฉลยข้อสอบ ส่วนในด้านต้องการเลือกข้อสอบไปใช้ส่วนใหญ่ต้องการเลือกตาม วิชา จุดประสงค์ ความยากง่าย อำนาจจำแนก ค่าการเดา และพฤติกรรมที่สอบ

ขั้นตอนที่ 2 พัฒนาระบบคลังข้อสอบ นำผลที่ได้จากการศึกษาในขั้นตอนที่ 1 และเอกสารงานวิจัยที่เกี่ยวข้อง มาทำการออกแบบระบบและเขียนโดยใช้ภาษาเดสก์ไพล์ได้ระบบคลังข้อสอบที่ประกอบไปด้วยระบบย่อย 6 ระบบ คือ ระบบรักษาความปลอดภัย ระบบการวางแผนสร้างคลังข้อสอบ ระบบจัดเก็บข้อสอบ ระบบสารสนเทศ ระบบคัดเลือกข้อสอบ และระบบการพิมพ์แบบสอบ หลังจากนั้นทำการเลือกภาษาและเขียนโปรแกรม ทำการทดสอบระบบ และจัดทำเอกสารคู่มือ ผลปรากฏว่าได้โปรแกรมระบบคลังข้อสอบที่ไม่มีความคลาดเคลื่อนทางภาษาคอมพิวเตอร์ และสามารถทำงานได้ตามวัตถุประสงค์ของการพัฒนาโปรแกรมระบบคลังข้อสอบทุกประการ

ขั้นตอนที่ 3 ประเมินผลระบบคลังข้อสอบ ทำการประเมินโดยการฝึกอบรมการพัฒนา ระบบคลังข้อสอบด้วยไมโครคอมพิวเตอร์ กลุ่มตัวอย่างเป็นอาจารย์ที่สนใจ จำนวน 25 คน ทำการฝึกอบรม 1 วัน และให้ผู้เข้าอบรมเป็นผู้ประเมิน ผลปรากฏว่าโปรแกรมระบบคลังข้อสอบที่พัฒนาขึ้นมีความเหมาะสมในการใช้ในโรงเรียนได้จริง โดยในด้านคู่มือการใช้โปรแกรมด้านการทำงาน ของโปรแกรม และการบริหารจัดการของโปรแกรมระบบคลังข้อสอบ มีความเหมาะสมมากที่สุด นำไปใช้งาน

จากการศึกษางานวิจัยทั้งหมดฉบับ มีประเด็นพิจารณา 3 ประเด็นคือ ประเด็นที่ 1 หน้าที่ การทำงานของคลังข้อสอบ ประเด็นที่ 2 ลักษณะแบบสอบ และ ประเด็นสุดท้าย ระบบปฏิบัติการ โปรแกรมคอมพิวเตอร์ที่ใช้ในการสร้างคลังข้อสอบ

ในประเด็นแรก หน้าที่การทำงานของคลังข้อสอบ จากงานวิจัยที่ศึกษา สามารถแบ่ง ลักษณะคลังข้อสอบออกเป็น 3 ประเภท ได้แก่ การทำหน้าที่ด้านการจัดเก็บข้อสอบ 3 งานวิจัย (ประคอง กรรณสูต, 2536; พลกร กรพิทักษ์, 2533; วีรยุทธ์ ธานี, 2546) การทำหน้าที่จัดเก็บและ วิเคราะห์ข้อสอบ จำนวน 2 งานวิจัย (กฤษฎาพันธ์ พงษ์บริบูรณ์, 2543; วินุลาศ เจริญชัย, 2537) และกลุ่มสุดท้าย คือ การจัดเก็บ ดำเนินการสอบ และวิเคราะห์ข้อสอบ (กอบกิจ สหัสรังษี, 2539)

ประเด็นที่สอง ลักษณะแบบสอบ จากงานวิจัยที่ศึกษา สามารถแบ่งลักษณะแบบสอบ 2 กลุ่ม ได้แก่ คลังข้อสอบที่จัดเก็บแบบสอบเลือกตอบเพียงอย่างเดียว 5 งานวิจัย (กฤษฎาพันธ์

พงษ์บริบูรณ์, 2543; ประคอง กรรณสูต, 2536; พลกร กรพิทักษ์, 2533; วินุลาศ เจริญชัย, 2537; วีรยุทธ์ ธานี, 2546) โดยงานวิจัยในกลุ่มนี้มีงานวิจัยของ งานวิจัยของ วีรยุทธ์ ธานี ที่สามารถจัดเก็บข้อสอบได้มากกว่าข้อความ โดยสามารถจัดเก็บข้อสอบที่เป็นรูปภาพ หรือสัญลักษณ์ทางคณิตศาสตร์ได้เพิ่มเติมกว่างานวิจัยทั่วไป สำหรับงานวิจัยของ กฤษฎาพันธ์ พงษ์บริบูรณ์ มีจุดเด่นในการดำเนินการสอบปฏิบัติการในคอมพิวเตอร์ได้อีกด้วย สำหรับงานวิจัยกลุ่มที่ 2 เป็นของ สามารถจัดเก็บข้อสอบ ได้ทั้ง แบบสอบเลือกตอบ แบบสอบถูกผิด แบบสอบจับคู่ และแบบสอบเติมคำ (กอบกิจ สหัสรังษี, 2539)

ประเด็นสุดท้าย ระบบปฏิบัติการโปรแกรมคอมพิวเตอร์ที่ใช้ในการสร้างคลังข้อสอบ แบ่งออกเป็น 2 กลุ่มคือ การใช้ระบบปฏิบัติการแบบ Dos จำนวน 5 งานวิจัย ภาษาที่นิยมใช้เขียนจะเป็นภาษา FoxPro ซึ่งจากงานวิจัยที่ผ่านมา สามารถเขียนภาษา ฟอรัทเรน Dbase, Foxbase ได้อีกด้วย (กฤษฎาพันธ์ พงษ์บริบูรณ์, 2543; กอบกิจ สหัสรังษี, 2539; ประคอง กรรณสูต, 2536; พลกร กรพิทักษ์, 2533; วินุลาศ เจริญชัย, 2537) มีงานวิจัยของ วีรยุทธ์ ธานี (2546) เพียงหนึ่งเรื่องที่ใช้ระบบปฏิบัติการ Windows พัฒนาค้างข้อสอบ ซึ่งโปรแกรม Microsoft windows นี้มีโปรแกรมเฉพาะการใช้งานให้เลือกสรรมากมาย เช่น Microsoft word, Microsoft powerpoint, Microsoft excel และ Microsoft access ซึ่งโปรแกรมเหล่านี้มีความเหมาะสมในการจัดทำฐานข้อมูล และยังไม่พบว่าม้งานวิจัยที่จัดทำคลังข้อสอบด้วยโปรแกรม Microsoft access ทั้งที่เป็นโปรแกรมที่มีอยู่แล้วสำหรับคอมพิวเตอร์ที่ใช้ระบบปฏิบัติการ Windows ซึ่งใช้แพร่หลายทั่วโลก ผู้วิจัยจึงมีความสนใจพัฒนาค้างข้อสอบด้วยโปรแกรม Microsoft access ครั้งนี้ขึ้นใช้ในการจัดเก็บคลังข้อสอบวิชาการวัดและการประเมินผลทางการศึกษา

กรอบแนวคิดการวิจัย

กรอบแนวคิดการวิจัยมี 2 ประการ คือ กรอบแนวคิดแรกเป็นการศึกษาและจำแนกคุณลักษณะของข้อสอบวิชาตามเกณฑ์พฤติกรรมการเรียนรู้ 6 ลักษณะ ประเภทของเนื้อหาวิชา 11 หัวข้อ และคุณภาพข้อสอบรายข้อด้านความยาก 5 ประเภท และด้านอำนาจจำแนก 5 ประเภท ตามแนวคิดการประเมินแบบอิงกลุ่ม ทฤษฎีการทดสอบแบบดั้งเดิม (Classical test theory) และกรอบแนวคิดของการเพื่อพัฒนาค้างข้อสอบวิชา โดยโปรแกรมคอมพิวเตอร์ ตามแนวคิดที่ประยุกต์จาก ศิริชัย กาญจนวาสี (2541)

กรอบแนวคิดแรก เป็นการศึกษาและจำแนกคุณลักษณะของข้อสอบวิชาตามเกณฑ์พฤติกรรมการเรียนรู้ ตามแนวคิดของ บลูม และคณะ (Bloom, B.S. et al., 1956) มี 6 ประเภท ประกอบด้วย ความรู้ความจำ ความเข้าใจ การประยุกต์ การวิเคราะห์ การสังเคราะห์ และการประเมิน ประเภทของเนื้อหาจำแนกตามประมวลรายวิชา 2702303 ปีการศึกษา 2546 ประกอบ

ด้วย 11 หัวข้อ ได้แก่ 1) หลักการวัดและการประเมินผล / การวัดประเมินตาม พระราชบัญญัติ การศึกษาแห่งชาติ พ.ศ. 2542 2) การประเมินการเรียนรู้ 3) เครื่องมือวัดผล 4) การสร้างเครื่องมือ วัดและประเมินผลการเรียนรู้ 5) การตรวจสอบคุณภาพเครื่องมือรายข้อ 6) การตรวจสอบคุณภาพ เครื่องมือทั้งฉบับ 7) ทฤษฎี IRT และการสร้างธนาคารข้อสอบ 8) การประเมินจิตพิสัย 9) การประเมินทักษะการปฏิบัติ 10) การตัดสินผลการเรียนรู้ และ 11) การประเมินคุณภาพภายใน และภายนอก และคุณภาพข้อสอบรายข้อด้านความยาก 5 ประเภท และด้านอำนาจจำแนก 5 ประเภท ตามแนวคิดการประเมินแบบอิงกลุ่ม ทฤษฎีการทดสอบแบบดั้งเดิม (Classical test theory) ซึ่งข้อสอบจะถูกวิเคราะห์คุณภาพรายข้อดังกล่าวด้วยโปรแกรม Classical test item analysis (CTIA Version 6.30) พัฒนาโดย สุพัฒน์ สุกมลสันต์ (2535) ซึ่งศูนย์ทดสอบทางการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ได้ดำเนินการวิเคราะห์และส่งรายงานผล การวิเคราะห์คุณภาพข้อสอบรายข้อหลังจากการสอบกลางภาคและปลายภาค โดยมีรายละเอียด ของการคำนวณค่าความยาก และอำนาจจำแนก ดังนี้

ความยาก (P: Difficulty indexes) หมายถึง สัดส่วนของจำนวนคนที่ตอบข้อสอบข้อนั้นได้ ถูกต้องต่อจำนวนผู้เข้าสอบทั้งหมด ดังนั้นหากข้อสอบที่ผู้สอบตอบข้อสอบได้เป็นจำนวนมาก ก็ จะถือว่าเป็นข้อที่ง่าย และข้อสอบที่ผู้สอบตอบได้จำนวนน้อย ก็ถือว่าเป็นข้อสอบที่มีความยาก

อำนาจจำแนก (r หรือ D: Discrimination indexes) หมายถึง ความสามารถของข้อสอบ ในการจำแนก/แยก ความแตกต่างระหว่างผู้สอบที่มีผลสัมฤทธิ์ทางการเรียนต่างกัน

ความยากและอำนาจจำแนก มีรายละเอียดของสูตรการคำนวณ และเกณฑ์การแปล ความหมาย ดังนี้ (ศิริชัย กาญจนวาสี, 2544; ดวงกมล ไตรวิจิตรคุณ, 2541)

สูตร

$$P = \frac{R_H + R_L}{N_H + N_L}$$

$$r = \frac{R_H - R_L}{N_H \text{ or } N_L}$$

เมื่อ	P	ค่าความยาก
	r	ค่าอำนาจจำแนก
	R_H	จำนวนคนที่ตอบถูกในกลุ่มสูง
	R_L	จำนวนคนที่ตอบถูกในกลุ่มต่ำ

N_H	จำนวนคนทั้งหมดในกลุ่มสูง
N_L	จำนวนคนทั้งหมดในกลุ่มต่ำ

เกณฑ์การพิจารณาคูณภาพข้อสอบ (ศิริชัย กาญจนวาสี, 2544; ดวงกมล ไตรวิจิตรคุณ, 2541)

ความยาก (p)	ความหมาย
0.00 – 0.19	ข้อสอบยาก ควรตัดทิ้งหรือ ปรับปรุง
0.20 – 0.39	ข้อสอบค่อนข้างยาก สามารถเก็บไว้ใช้ได้ต่อไป
0.40 – 0.60	ข้อสอบยากพอเหมาะ สามารถเก็บไว้ใช้ได้ต่อไป
0.61 – 0.80	ข้อสอบค่อนข้างง่าย สามารถเก็บไว้ใช้ได้ต่อไป
0.81 – 1.00	ข้อสอบง่าย ควรตัดทิ้ง หรือ ปรับปรุง

อำนาจจำแนก (r)	ความหมาย
ต่ำกว่า 0	ข้อสอบจำแนกไม่ได้ ควรตัดทิ้ง หรือ ปรับปรุง
0.00 – 0.19	ข้อสอบจำแนกได้ต่ำ ควรตัดทิ้ง หรือ ปรับปรุง
0.20 – 0.29	ข้อสอบจำแนกได้พอใช้ สามารถเก็บไว้ใช้ได้ต่อไป
0.30 – 0.39	ข้อสอบจำแนกได้ดี สามารถเก็บไว้ใช้ได้ต่อไป
0.40 ขึ้นไป	ข้อสอบจำแนกได้ดีมาก สามารถเก็บไว้ใช้ได้ต่อไป

ดังนั้นข้อสอบที่สามารถเก็บไว้ใช้ได้ต่อไป จึงควรมีค่าความยาก อยู่ระหว่าง 0.20 – 0.80 และค่าอำนาจจำแนก ตั้งแต่ 0.20 ขึ้นไป ซึ่งเป็นการคัดเลือกข้อสอบตามแนวคิดการประเมินแบบอิงกลุ่ม สาเหตุที่ไม่ได้คัดเลือกข้อสอบเก็บเข้าคลังข้อสอบตามแนวการประเมินแบบอิงเกณฑ์ เนื่องจากในเชิงปฏิบัติการสอบกลางภาคและปลายภาคเป็นการจัดสอบครั้งเดียวกับกลุ่มผู้เรียนซึ่งไม่เหมาะสมสำหรับการคำนวณ ค่าความไว ซึ่งเป็นดัชนีบ่งชี้ความสามารถในการจำแนกบุคคลระหว่างกลุ่มรอบรู้และกลุ่มไม่รอบรู้ หรือผู้ที่เรียนในเนื้อหานั้นมาแล้วและผู้ยังที่ไม่ได้เรียน ตามแนวคิดการประเมินแบบอิงเกณฑ์ได้

ดังนั้นผู้วิจัยจึงเลือกใช้แนวการประเมินแบบอิงกลุ่มเป็นเกณฑ์การคัดเลือกข้อสอบ ซึ่งค่าความยากที่คัดเลือกระหว่าง 0.2 – 0.8 แสดงว่ามีผู้สอบจำนวนร้อยละ 20 ถึง 80 ทำข้อนั้นได้ถูกต้อง ถือว่ามีความยากง่ายปานกลาง ไม่เป็นข้อสอบที่ง่ายมากหรือยากมากเกินไป สำหรับค่าอำนาจจำแนกที่ใช้ คัดเลือกข้อสอบมีค่ามากกว่าเท่ากับ 0.2 แสดงว่า กลุ่มที่ทำคะแนนได้สูงมีจำนวนผู้ตอบถูกต้องมากกว่ากลุ่มต่ำ ค่าอำนาจจำแนกมีค่าบวก ซึ่งเป็นลักษณะที่ตรงกันข้ามกับ

คำอำนาจจำแนกติดลบ ที่เป็นผลจากจำนวนคนในกลุ่มต่ำทำถูกต้องมากกว่ากลุ่มสูง ซึ่งเป็นเครื่องบ่งชี้ว่า อาจมีความกำกวมของคำถาม หรือเฉลยผิดเกิดขึ้นได้

กรอบแนวคิดสุดท้าย คือ การเพื่อพัฒนาค้างข้อสอบวิชา โดยโปรแกรมคอมพิวเตอร์ที่ประยุกต์แนวคิดจาก ศิริชัย กาญจนวาสี (2541) ซึ่งเสนอแนะขั้นตอนการพัฒนาระบบคลังข้อสอบด้วยโปรแกรมคอมพิวเตอร์ ประกอบด้วย 7 ขั้นตอน ได้แก่ การวางแผน การออกแบบโปรแกรม การจัดเก็บข้อสอบเข้าคลัง การคัดเลือกข้อสอบจากคลังมาใช้งาน การจัดพิมพ์ข้อสอบ การให้สารสนเทศเกี่ยวกับข้อสอบ การรักษาความปลอดภัยของระบบคลังข้อสอบเป็นกรอบสำหรับการพัฒนา และเนื่องจากคลังข้อสอบที่สร้างขึ้นไม่สามารถตรวจให้คะแนนและวิเคราะห์ข้อสอบได้ จึงไม่มีขั้นตอนของการให้สารสนเทศเกี่ยวกับแบบสอบ อย่างไรก็ตามผู้วิจัยได้เพิ่มขั้นตอน การจำแนกข้อสอบตามลักษณะเฉพาะ การคัดเลือกข้อสอบ และการเขียนโปรแกรมเพิ่มขึ้นอีก 3 ขั้นตอน รวมเป็น 9 ขั้นตอน



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 3

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้ดำเนินการใช้ระเบียบวิธีวิจัยเชิงบรรยาย (Descriptive research) และการวิจัยเชิงพัฒนา (Developmental research) โดยมุ่งบรรยายและจำแนกคุณลักษณะของข้อสอบวิชา 2702303 การวัดและการประเมินผลทางการศึกษาตามเกณฑ์พฤติกรรมการเรียนรู้ เนื้อหาวิชา และคุณภาพของข้อสอบรายข้อด้านความยากและอำนาจจำแนก พร้อมทั้งพัฒนาค้างข้อสอบโดยโปรแกรมคอมพิวเตอร์เพื่อให้เกิดรวบรวมข้อสอบที่มีคุณภาพ สำหรับคัดเลือกไว้ในโอกาสต่อไป

วิธีดำเนินการวิจัยการพัฒนาค้างข้อสอบวิชา 2702303 การวัดและการประเมินผลทางการศึกษา ด้วยโปรแกรมคอมพิวเตอร์นี้ ประกอบด้วยหัวข้อ ประชากรข้อสอบ การพัฒนาโปรแกรมคอมพิวเตอร์ และการวิเคราะห์ข้อมูลจำแนกประเภทของข้อสอบตามวัตถุประสงค์ข้างต้น

ข้อสอบที่นำมาวิเคราะห์

ข้อสอบที่นำมาศึกษาประกอบด้วย ข้อสอบในรายวิชา 2702303 การวัดและการประเมินผลทางการศึกษา ปีการศึกษา 2545-2546 จำนวน 260 ข้อ และข้อสอบในวิชา 2702302 การประเมินผลการเรียนการสอน เฉพาะที่มีเนื้อหาเกี่ยวข้องกับวิชา 2702303 และผ่านการวิเคราะห์ค่าความยากและค่าอำนาจจำแนกแล้ว ประกอบด้วย ข้อสอบในปีการศึกษา 2541 จำนวน 207 ข้อ และปีการศึกษา 2544 จำนวน 130 ข้อ ผู้วิจัยคัดเลือกข้อสอบทั้งหมดมาศึกษาจำนวน 597 ข้อ

การพัฒนาโปรแกรมคอมพิวเตอร์

โปรแกรมคอมพิวเตอร์ที่ใช้พัฒนาค้างข้อสอบวิชา 2702303 การวัดประเมินผลการศึกษา คือ โปรแกรมไมโครซอฟต์แอสเซส (Microsoft access) โดยผู้วิจัย และพนัส จันทร์เปล่ง ได้ร่วมกันวางรูปแบบและพัฒนาโปรแกรมค้างข้อสอบนี้ขึ้น เพื่อเก็บรวบรวมข้อสอบที่มีคุณภาพสำหรับใช้ในโอกาสต่อไป โดยมีวิธีดำเนินการดังนี้

1. การวางแผนและกำหนดลักษณะค้างข้อสอบ ประกอบด้วย
 - 1.1 ลำดับของข้อคำถาม
 - 1.2 ปีการศึกษา
 - 1.3 ภาคเรียน

- 1.4 ข้อคำถาม/โจทย์
 - 1.5 คำตอบหรือตัวเลือก 5 ตัวเลือก
 - 1.6 ค่าความยาก
 - 1.7 คำอธิบายแปลผลค่าความยาก
 - 1.8 คำอำนาจจำแนก
 - 1.9 คำอธิบายแปลผลอำนาจจำแนก
 - 1.10 พฤติกรรมการเรียนรู้
 - 1.11 เนื้อหารายวิชา
2. การจำแนกข้อสอบตามลักษณะตัวแปรในข้อที่ 1
 3. การคัดเลือกข้อสอบที่มีคุณภาพ ค่าความยากอยู่ในช่วง 0.2-0.8 และคำอำนาจจำแนกตั้งแต่ 0.2 ขึ้นไปเพื่อจัดเก็บในโปรแกรม
 4. การออกแบบโปรแกรมคอมพิวเตอร์โดยเลือกโปรแกรม Microsoft access
 5. การเขียนคำสั่งโปรแกรม Microsoft access
 6. การจัดเก็บข้อสอบในคลังข้อสอบ
 7. การคัดเลือกข้อสอบมาใช้งาน โดยกำหนดเลือกข้อสอบจาก ปีการศึกษา ภาคเรียน ค่าความยาก อำนาจจำแนก พฤติกรรมการเรียนรู้ และเนื้อหาวิชา
 8. การจัดพิมพ์ข้อสอบโดยมีให้เลือกจัดพิมพ์ 2 ลักษณะคือ
 - 8.1 จัดพิมพ์ให้ปรากฏข้อคำถาม ตัวเลือก ค่าความยาก คำอธิบายแปลผลค่าความยาก คำอำนาจจำแนก คำอธิบายแปลผลค่าอำนาจจำแนก พฤติกรรมการเรียนรู้ และ เนื้อหาวิชา
 - 8.2 จัดพิมพ์ในโปรแกรม Microsoft word
 9. การรักษาความปลอดภัยของข้อสอบโดยกำหนดรหัส Password ในการเปิดโปรแกรมใช้งาน

การเก็บรวบรวมข้อมูล

1. ศึกษาแบบสอบกลางภาคและปลายภาคตั้งแต่ปีการศึกษา 2537 – 2546 ในรายวิชา 2702302 และ 2702303 ที่จัดเก็บไว้ พบว่ามีผลการวิเคราะห์ข้อสอบด้วยค่าความยาก และ คำอำนาจจำแนกเฉพาะในปีการศึกษา 2541 ปีการศึกษา 2544-2546 เท่านั้น
2. คัดเลือกข้อสอบที่มีข้อมูลค่าความยากและคำอำนาจจำแนก มาจำแนกข้อสอบตามลักษณะพฤติกรรมการเรียนรู้ และเนื้อหาวิชา จำนวน 597 ข้อ ซึ่งมาจากต่างรายวิชา มีแผนผัง

ข้อสอบ และจำนวนข้อสอบแตกต่างกัน การดำเนินงานวิจัยนี้ให้ความสำคัญกับการพัฒนาคลังข้อสอบเพื่อจัดเก็บข้อสอบที่มีคุณภาพจัดเก็บไว้ใช้ในการสร้างแบบสอบปีการศึกษาต่อไปเท่านั้น

3. คัดเลือกเฉพาะข้อสอบที่มีค่าความยากอยู่ในช่วง 0.2-0.8 และค่าอำนาจจำแนกตั้งแต่ 0.2 ขึ้นไปจำนวน 322 ข้อ เพื่อจัดเก็บไว้ในคลังข้อสอบ

4. เนื่องจากโปรแกรมคลังข้อสอบที่พัฒนาขึ้นจัดเก็บได้เฉพาะข้อที่มีโจทย์คำถามเป็นข้อความเท่านั้น ไม่สามารถจัดเก็บข้อสอบที่มีโจทย์เป็นตารางได้ ซึ่งต้องมีการพัฒนาโปรแกรมต่อไป จึงไม่สามารถเก็บข้อสอบได้ครบทั้ง 322 ข้อ

การวิเคราะห์ข้อมูล

จำแนกคุณลักษณะของข้อสอบ ตามเกณฑ์พฤติกรรมการเรียนรู้ ประเภทของเนื้อหา ด้วยสถิติความถี่ ร้อยละ และวิเคราะห์คุณภาพข้อสอบรายข้อด้านความยากและอำนาจจำแนก ด้วยการคำนวณค่าเฉลี่ยเลขคณิต ส่วนเบี่ยงเบนมาตรฐาน ค่าความเบ้ ค่าความโด่ง เป็นต้น



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 4

ผลการวิเคราะห์ข้อมูล

การวิจัยนี้มีวัตถุประสงค์หลักเพื่อศึกษาและจำแนกคุณลักษณะของข้อสอบวิชา 2702303 การวัดและการประเมินผลทางการศึกษาตามเกณฑ์พฤติกรรมการเรียนรู้ ประเภทของเนื้อหาวิชา และคุณภาพข้อสอบรายข้อด้านความยากและอำนาจจำแนก และเพื่อพัฒนาคลังข้อสอบวิชา 2702303 การวัดและการประเมินผลทางการศึกษาโดยโปรแกรมคอมพิวเตอร์ เพื่อเก็บรวบรวมข้อสอบที่มีคุณภาพไว้ใช้ในโอกาสต่อไป

ผลการวิเคราะห์ข้อมูลที่น่าเสนอในบทนี้แบ่งการนำเสนอเป็น 4 ตอน ดังต่อไปนี้

ตอนที่ 1 ผลการวิเคราะห์ข้อสอบวิชา 2702303 การวัดและการประเมินผลทางการศึกษา

ตอนที่ 2 ผลการวิเคราะห์ข้อสอบที่จะคัดเลือกไว้ในคลังข้อสอบ

ตอนที่ 3 ผลการวิเคราะห์ค่าความเที่ยงของแบบสอบ

ตอนที่ 4 ลักษณะคลังข้อสอบวิชา 2702303 การวัดและการประเมินผลการศึกษา

ตอนที่ 1 ผลการวิเคราะห์ข้อสอบวิชา 2702303 การวัดและการประเมินผลการศึกษา

ข้อสอบวิชา 2702303 การวัดและการประเมินผลการศึกษา ที่มีผลการวิเคราะห์คุณภาพข้อสอบส่วนใหญ่เป็นข้อสอบในปีการศึกษา 2541 จำนวน 207 ข้อ หรือ ร้อยละ 34.7 รองลงมา คือ ข้อสอบในปีการศึกษา 2544-2546 จำนวนเท่า ๆ กัน 130 ข้อ หรือ ร้อยละ 21.8 สำหรับข้อสอบปีการศึกษา 2542-2543 ขาดการการจัดเก็บคุณภาพข้อสอบไว้จึงไม่ได้นำมาศึกษา

จากข้อสอบจำนวนทั้งสิ้น 597 ข้อ เป็นการวัดพฤติกรรมกรเรียนรู้ในด้านความเข้าใจมากที่สุด ร้อยละ 31 รองลงมาเป็นความรู้ความจำ ร้อยละ 28.1 การวิเคราะห์ ร้อยละ 26 การนำไปใช้ ร้อยละ 10 การสังเคราะห์ ร้อยละ 3 สำหรับการวัดพฤติกรรมกรเรียนรู้ด้านการประเมินมีจำนวนน้อยที่สุด ร้อยละ 1.8

เนื้อหารายวิชาที่มีจำนวนข้อสอบมากที่สุดคือ หลักการวัดและการประเมินผล/การวัด ประเมินผลตามพระราชบัญญัติการศึกษา 2542 ร้อยละ 15.9 รองลงมาคือ การตรวจสอบคุณภาพเครื่องมือทั้งฉบับ ร้อยละ 13.7 และ การตรวจสอบคุณภาพเครื่องมือรายข้อ ร้อยละ 12.7 เนื้อหามีข้อสอบวัดน้อยที่สุดคือ ทฤษฎี IRT และการสร้างธนาคารข้อสอบ ร้อยละ 3.9 และรองลงมา คือ

การประเมินคุณภาพภายในและภายนอก ร้อยละ 4.5 ซึ่งจัดเป็นเนื้อหาใหม่ที่เริ่มสอนเมื่อปีการศึกษา 2545-2546 ดังรายละเอียดในตารางที่ 1

ตารางที่ 1 ความดีร้อยละของข้อสอบจำแนกตามตัวแปร

ลักษณะข้อสอบ	ความถี่	ร้อยละ
1. ปีการศึกษา		
2541	207	34.7
2544	130	21.8
2545	130	21.8
2546	130	21.8
2. พฤติกรรมการเรียนรู้		
ความรู้ความจำ	168	28.1
ความเข้าใจ	185	31.0
การนำไปใช้	60	10.1
การวิเคราะห์	155	26.0
การสังเคราะห์	18	3.0
การประเมิน	11	1.8
3. เนื้อหา		
หลักการวัดและการประเมินผล / การวัดประเมินตามพรบ.	95	15.9
การประเมินการเรียนรู้	39	6.5
เครื่องมือวัดผล	67	11.2
การสร้างเครื่องมือวัดและประเมินผลการเรียนรู้	55	9.2
การตรวจสอบคุณภาพเครื่องมือรายข้อ	73	12.7
การตรวจสอบคุณภาพเครื่องมือทั้งฉบับ	82	13.7
ทฤษฎี IRT และการสร้างธนาคารข้อสอบ	23	3.9
การประเมินจิตพิสัย	34	5.7
การประเมินทักษะการปฏิบัติ	39	6.5
การตัดสินผลการเรียนรู้	60	10.1
การประเมินคุณภาพภายในและภายนอก	27	4.5
รวม	597	100

เมื่อวิเคราะห์ข้อสอบจำแนกตามเนื้อหาวิชาและพฤติกรรมการเรียนรู้ พบว่า หลักการวัดและการประเมินผลตามพระราชบัญญัติการศึกษา 2542 มีข้อสอบที่วัดพฤติกรรมการเรียนรู้ด้านความรู้ มากที่สุดร้อยละ 37.9 รองลงมาคือ ความเข้าใจ ร้อยละ 33.7 เนื้อหาการประเมินผล การเรียนรู้ พฤติกรรมการเรียนที่วัดมากที่สุดคือ ด้านความเข้าใจ ร้อยละ 38.5 รองลงมา คือ ด้านการวิเคราะห์ ร้อยละ 35.9 เนื้อหาเครื่องมือวัดผล พฤติกรรมการเรียนรู้ที่วัดมากที่สุด คือ ด้านความรู้ ร้อยละ 50.7 รองลงมาคือ ด้านการนำไปใช้ ร้อยละ 20.9 เนื้อหาการสร้างเครื่องมือวัดและประเมินผลการเรียนรู้ พฤติกรรมการเรียนรู้ที่วัดมากที่สุดคือ ด้านความเข้าใจ ร้อยละ 60 รองลงมาคือ ด้านความรู้ร้อยละ 14.5 เนื้อหาการตรวจสอบคุณภาพของเครื่องมือรายข้อจาก ทฤษฎีการทดสอบแบบดั้งเดิม พฤติกรรมการเรียนรู้ที่วัดมากที่สุดคือ ด้านความเข้าใจ ร้อยละ 43.4 รองลงมาคือ ด้านการวิเคราะห์ ร้อยละ 32.9 เนื้อหาการตรวจสอบคุณภาพเครื่องมือทั้งฉบับ พฤติกรรมการเรียนรู้ที่วัดมากที่สุดคือ ด้านการวิเคราะห์ ร้อยละ 37.8 รองลงมาคือ ด้านความเข้าใจ ร้อยละ 30.5 เนื้อหาทฤษฎีการตอบสนองข้อสอบและการสร้างธนาคารข้อสอบ พฤติกรรมการเรียนรู้ที่วัดมากที่สุดคือ ด้านความรู้ ร้อยละ 52.2 รองลงมาคือ ด้านความเข้าใจ ร้อยละ 39.1 เนื้อหา การประเมินจิตพิสัย พฤติกรรมการเรียนรู้ที่วัดมากที่สุดคือ ด้านความรู้ ร้อยละ 50 รองลงมา คือ ด้านการวิเคราะห์ ร้อยละ 20.6 เนื้อหาการประเมินทักษะการปฏิบัติ พฤติกรรมการเรียนรู้ที่วัดมากที่สุดคือ ด้านความรู้ ร้อยละ 35.9 รองลงมาคือ ด้านความเข้าใจและการนำไปใช้ ร้อยละ 20.5 เนื้อหาการตัดสินผลการเรียนรู้หรือการตัดสินการเรียนรู้ พฤติกรรมการเรียนรู้ที่วัดมากที่สุด คือ ด้านการวิเคราะห์ ร้อยละ 61.7 รองลงมาคือ ด้านความเข้าใจ ร้อยละ 23.3 เนื้อหาการประเมินคุณภาพภายในและภายนอก พฤติกรรมการเรียนรู้ที่วัดมากที่สุดคือ ด้านความรู้ ร้อยละ 66.7 รองลงมาคือ ด้านความเข้าใจ ร้อยละ 18.5

เมื่อพิจารณาจำแนกตามพฤติกรรมการเรียนรู้พบว่า การวัดระดับความรู้ ใช้มากที่สุด ในเนื้อหาหลักการวัดประเมินผล การวัดประเมินผลตามพระราชบัญญัติการศึกษา ร้อยละ 21.4 ด้านความเข้าใจ ใช้วัดมากที่สุด เนื้อหา การสร้างเครื่องมือ และการตรวจสอบคุณภาพเครื่องมือ รายข้อ ร้อยละ 17.8 ด้านการนำไปใช้ ใช้วัดมากที่สุด เนื้อหา การตรวจสอบคุณภาพข้อสอบ รายข้อและเครื่องมือทั้งฉบับ ร้อยละ 13.3 ด้านการวิเคราะห์ ใช้วัดมากที่สุด เนื้อหาการตัดสินผล การเรียนรู้ ร้อยละ 23.9 ด้านการสังเคราะห์ ใช้วัดมากที่สุด เนื้อหาเครื่องมือวัดผล การตรวจสอบคุณภาพข้อสอบรายข้อและเครื่องมือทั้งฉบับ ร้อยละ 22.2 ด้านการประเมินผล ใช้วัดมากที่สุด ในเนื้อหาหลักการวัดประเมินผล การวัดประเมินผลตามพระราชบัญญัติการศึกษา ร้อยละ 45.5 ดังแสดงรายละเอียดในตารางที่ 2

ตารางที่ 3 ค่าเฉลี่ยเลขคณิต ส่วนเบี่ยงเบนมาตรฐาน ความเบ้ และความโด่งของค่าความยากและอำนาจจำแนกของข้อสอบ

ตัวแปร	MIN	MAX	MEAN	S.D.	SK	KU
ความยาก	0.033	1.00	0.63	0.21	-0.52	-0.36
อำนาจจำแนก	-0.38	0.67	0.26	0.16	-0.08	-0.29

ข้อสอบที่นำมาวิเคราะห์คุณภาพรายข้อ ด้านความยาก อำนาจจำแนก พบว่า ข้อสอบมีค่าความยากเฉลี่ย เท่ากับ 0.63 ซึ่งถือว่าข้อสอบค่อนข้างง่าย ส่วนเบี่ยงเบนมาตรฐานมีค่าเท่ากับ 0.21 มีลักษณะเบ้ซ้าย คือข้อสอบส่วนใหญ่ค่อนข้างง่าย ความโด่งมีลักษณะแบน

ค่าอำนาจจำแนกเฉลี่ย เท่ากับ 0.26 ซึ่งเป็นค่าอำนาจจำแนกได้พอใช้ ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.16 มีลักษณะเบ้ซ้าย และความโด่งมีลักษณะแบน

ตารางที่ 4 ค่าความถี่และร้อยละของคุณภาพข้อสอบจำแนกตามความยากและการนำไปใช้

คุณภาพข้อสอบ	ความถี่	ร้อยละ
1. ความยาก		
ข้อสอบยาก ควรตัดทิ้งหรือปรับปรุง	17	2.8
ข้อสอบค่อนข้างยาก สามารถเก็บไว้ใช้ได้ต่อไป	73	12.2
ข้อสอบยากพอเหมาะ สามารถเก็บไว้ใช้ได้ต่อไป	153	25.6
ข้อสอบค่อนข้างง่าย สามารถเก็บไว้ใช้ได้ต่อไป	220	36.9
ข้อสอบง่าย ควรตัดทิ้งหรือปรับปรุง	134	22.4
2. การนำข้อสอบไปใช้ โดยพิจารณาความยาก		
ควรตัดทิ้งหรือปรับปรุง	151	25.3
สามารถเก็บไว้ใช้ได้ต่อไป	446	74.7
รวม	597	100

จากคุณภาพความยากของข้อสอบทั้งหมดส่วนใหญ่เป็นข้อสอบค่อนข้างง่าย มีจำนวนมากที่สุด ร้อยละ 36.9 รองลงมาคือ ข้อสอบยากพอเหมาะ ร้อยละ 25.6 ข้อสอบที่สามารถเก็บไว้ใช้ได้ต่อไป คือ ข้อสอบค่อนข้างง่าย ยากพอเหมาะ และค่อนข้างยาก ซึ่งมีค่าความยากอยู่ในช่วง 0.2 – 0.8 มีจำนวนทั้งสิ้น 446 ข้อ คิดเป็นร้อยละ 74.7

ตารางที่ 5 ค่าความถี่และร้อยละของคุณภาพข้อสอบจำแนกตามอำนาจจำแนกและการนำไปใช้

คุณภาพข้อสอบ	ความถี่	ร้อยละ
1. อำนาจจำแนก		
ข้อสอบจำแนกไปไม่ได้ ควรตัดทิ้งหรือปรับปรุง	24	4.0
ข้อสอบจำแนกได้ต่ำ ควรตัดทิ้งหรือปรับปรุง	188	31.5
ข้อสอบจำแนกได้พอใช้ สามารถเก็บไว้ใช้ได้ต่อไป	140	23.5
ข้อสอบจำแนกได้ดี สามารถเก็บไว้ใช้ได้ต่อไป	104	17.4
ข้อสอบจำแนกได้ดีมาก สามารถเก็บไว้ใช้ได้ต่อไป	141	23.6
2. การนำข้อสอบไปใช้ โดยพิจารณาอำนาจจำแนก		
ควรตัดทิ้งหรือปรับปรุง	212	35.5
สามารถเก็บไว้ใช้ได้ต่อไป	385	64.5
รวม	597	100

จากคุณภาพด้านอำนาจจำแนกของข้อสอบทั้งหมดส่วนใหญ่เป็นข้อสอบจำแนกได้ต่ำมีจำนวนมากที่สุด ร้อยละ 31.5 รองลงมาคือ ข้อสอบจำแนกพอใช้ ร้อยละ 23.5 ข้อสอบที่สามารถเก็บไว้ใช้ได้ต่อไป คือ ข้อสอบที่มีอำนาจจำแนกได้พอใช้ จำแนกได้ดี และจำแนกได้ดีมา ซึ่งมีค่าอำนาจจำแนกตั้งแต่ 0.2 เป็นต้นไป จำนวนทั้งสิ้น 386 ข้อ คิดเป็นร้อยละ 64.5

ตารางที่ 6 ความถี่และร้อยละ ของข้อสอบที่มีคุณภาพด้านความยากและอำนาจจำแนก

คุณภาพข้อสอบที่ผ่านเกณฑ์ความยากและอำนาจจำแนก	ความถี่	ร้อยละ
ควรตัดทิ้งหรือปรับปรุง	275	46.1
สามารถเก็บไว้ใช้ได้ต่อไป	322	53.9
รวม	597	100

เมื่อพิจารณาคุณภาพข้อสอบที่ผ่านเกณฑ์ความยากซึ่งมีค่าอยู่ในช่วง 0.2 – 0.8 และค่าอำนาจจำแนก ตั้งแต่ 0.2 เป็นต้นไป ถือเป็นข้อสอบที่มีคุณภาพและเก็บไว้ใช้ได้ต่อไป มีจำนวนทั้งสิ้น 322 ข้อ คิดเป็นร้อยละ 53.9

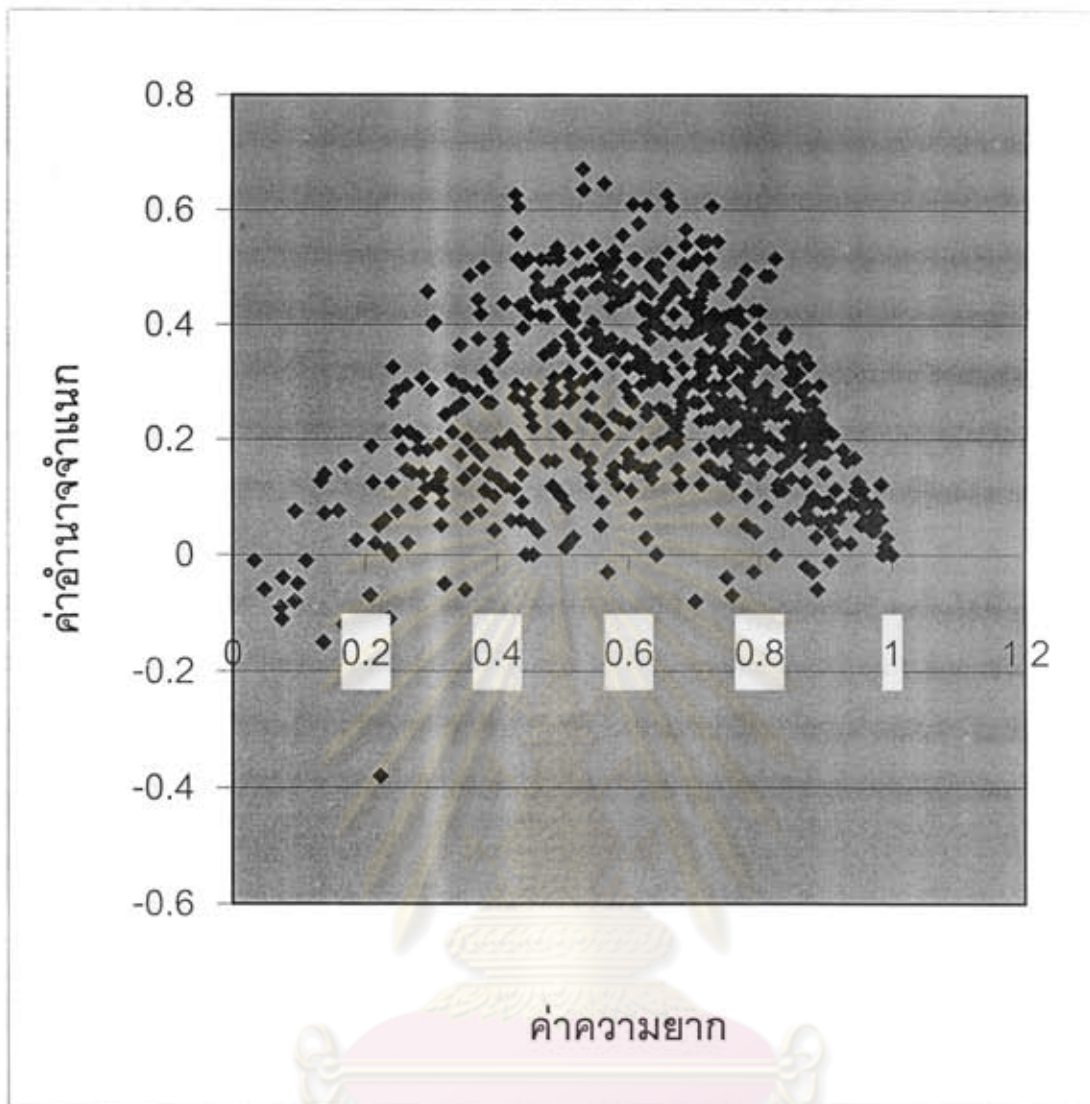
ตารางที่ 7 ความถี่และร้อยละของคุณภาพข้อสอบจำแนกตามความยากและอำนาจจำแนก

อำนาจจำแนก	ค่าความยาก		รวม
	ควรตัดทิ้งหรือปรับปรุง	สามารถเก็บไว้ใช้ต่อไป	
ควรตัดทิ้งหรือปรับปรุง	88 14.7%	124 20.8%	212 35.5%
สามารถเก็บไว้ใช้ต่อไป	63 10.6%	322 53.9%	385 64.5%
รวม	151 25.3%	446 74.7%	597 100.0%

การแจกแจงคุณภาพข้อสอบจากค่าความยากและอำนาจจำแนก พบว่า ข้อสอบผ่านเกณฑ์ค่าความยากและอำนาจจำแนกที่สามารถเก็บไว้ใช้ต่อไปได้มีจำนวน 322 ข้อ ร้อยละ 53.9 รองลงมาเป็นข้อสอบที่มีผ่านเกณฑ์ค่าความยากแต่ค่าอำนาจจำแนกควรตัดทิ้งหรือปรับปรุงมีจำนวน 124 ข้อ ร้อยละ 20.8 ข้อสอบที่ไม่ผ่านเกณฑ์ค่าความยากแต่มีค่าอำนาจจำแนกเหมาะสมมีจำนวน 63 ข้อ ร้อยละ 10.6 และ ข้อสอบที่ไม่ผ่านเกณฑ์ทั้งค่าความยากและอำนาจจำแนกมีจำนวน 88 ข้อ ร้อยละ 14.7

เมื่อพิจารณาแผนภูมิการกระจายค่าความยากและอำนาจจำแนก พบว่า ข้อสอบที่ง่ายมากหรือยากมากจะมีอำนาจจำแนกอยู่ในช่วงจำกัดและต่ำ ข้อสอบที่มีค่าอำนาจจำแนกตั้งแต่ 2 ขึ้นไป ส่วนใหญ่เป็นข้อสอบที่มีความยากปานกลางค่อนข้างง่าย ดังแสดงในภาพที่ 1

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย



ภาพที่ 1 การกระจายค่าความยากและอำนาจจำแนกของข้อสอบ

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตอนที่ 2 ผลการวิเคราะห์ข้อสอบที่จะคัดเลือกไว้ในคลังข้อสอบ

จากข้อสอบจำนวนทั้งสิ้น 597 ข้อ เป็นข้อสอบที่ผ่านเกณฑ์ค่าความยากและอำนาจจำแนกสามารถเก็บไว้ใช้ต่อไปได้จำนวน 322 ข้อ ข้อสอบเหล่านี้ เป็นข้อสอบจากปีการศึกษา 2541 มากที่สุด 110 ข้อ ร้อยละ 34.2 รองลงมาคือ ปีการศึกษา 2545 จำนวน 72 ร้อยละ 22.4 และปีการศึกษา 2544 ปีการศึกษา 2546 ในจำนวนเท่ากันคือ 70 ข้อ ร้อยละ 21.7

ข้อสอบที่เก็บไว้ใช้ เป็นข้อสอบที่วัดพฤติกรรมการเรียนรู้ในด้านความเข้าใจมากที่สุด ร้อยละ 32.3 รองลงมาเป็นความรู้ความจำ ร้อยละ 29.2 การวิเคราะห์ ร้อยละ 26.1 การนำไปใช้ ร้อยละ 7.1 การสังเคราะห์ ร้อยละ 3.4 สำหรับการวัดพฤติกรรมการเรียนรู้ด้านการประเมินมีจำนวนน้อยที่สุด ร้อยละ 1.9

เนื้อหารายวิชาที่มีจำนวนข้อสอบที่มีคุณภาพใช้ได้มากที่สุด คือ การวิเคราะห์ทั้งฉบับ ร้อยละ 19.6 รองลงมาคือ การวิเคราะห์คุณภาพรายข้อ ร้อยละ 14.9 และหลักการวัดและการประเมินผล/การวัดประเมินผลตามพระราชบัญญัติการศึกษา 2542 ร้อยละ 13.7 เนื้อหามีข้อสอบที่มีคุณภาพจำนวนน้อยที่สุดคือ การประเมินทักษะการปฏิบัติ ร้อยละ 3.1 ดังรายละเอียดในตารางที่ 8

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 8 ความถี่ร้อยละของข้อสอบที่สามารถเก็บไว้ใช้ได้ จำแนกตามตัวแปร

ลักษณะข้อสอบ	ทั้งหมด		เก็บไว้ใช้ได้	
	ความถี่	ร้อยละ	ความถี่	ร้อยละ
1. ปีการศึกษา				
2541	207	34.7	110	34.2
2544	130	21.8	70	21.7
2545	130	21.8	72	22.4
2546	130	21.8	70	21.7
2. พฤติกรรมการเรียนรู้				
ความรู้ความจำ	168	28.1	94	29.2
ความเข้าใจ	185	31.0	104	32.3
การนำไปใช้	60	10.1	23	7.1
การวิเคราะห์	155	26.0	84	26.1
การสังเคราะห์	18	3.0	11	3.4
การประเมิน	11	1.8	6	1.9
3. เนื้อหา				
หลักการวัดและการประเมินผล / การวัดประเมินตาม พรบ.	95	15.9	44	13.7
การประเมินการเรียนรู้	39	6.5	17	5.3
เครื่องมือวัดผล	67	11.2	30	9.3
การสร้างเครื่องมือวัดและประเมินผลการเรียนรู้	55	9.2	27	8.4
การตรวจสอบคุณภาพเครื่องมือรายข้อ	73	12.7	48	14.9
การตรวจสอบคุณภาพเครื่องมือทั้งฉบับ	82	13.7	63	19.6
ทฤษฎี IRT และการสร้างธนาคารข้อสอบ	23	3.9	13	4.0
การประเมินจิตพิสัย	34	5.7	20	6.2
การประเมินทักษะการปฏิบัติ	39	6.5	10	3.1
การตัดสินผลการเรียนรู้	60	10.1	36	11.2
การประเมินคุณภาพภายในและภายนอก	27	4.5	14	4.3
รวม	597	100	322	100

ผลการจำแนกข้อสอบตามเนื้อหาวิชาและพฤติกรรมการเรียนรู้พบว่า เนื้อหาวิชาหลักการวัดประเมินผล/การวัดประเมินผลตาม พ.ร.บ. พ.ศ. 2542 มีการวัดพฤติกรรมความรู้มากที่สุด ไม่ปรากฏข้อที่วัดพฤติกรรมด้านการสังเคราะห์

เนื้อหาการประเมินผลการเรียนรู้ วัดพฤติกรรมความเข้าใจมากที่สุด ไม่ปรากฏข้อสอบที่วัดพฤติกรรมด้านการประเมิน

เนื้อหาเครื่องมือวัดผล วัดพฤติกรรมความรู้มากที่สุด ไม่ปรากฏข้อสอบที่วัดพฤติกรรมสังเคราะห์และการประเมิน

เนื้อหาการสร้างเครื่องมือวัดพฤติกรรมความเข้าใจมากที่สุด ไม่ปรากฏข้อสอบที่วัดพฤติกรรมการสังเคราะห์

เนื้อหาคุณภาพข้อสอบรายข้อ วัดพฤติกรรมความเข้าใจมากที่สุด และมีข้อสอบที่วัดพฤติกรรมด้านความรู้และการประเมินจำนวนน้อยที่สุด

เนื้อหาคุณภาพข้อสอบทั้งฉบับ วัดพฤติกรรมวิเคราะห์มากที่สุด ไม่ปรากฏข้อสอบที่วัดพฤติกรรมการประเมิน

เนื้อหาทฤษฎีการตอบสนองข้อสอบและธนาคารข้อสอบ วัดพฤติกรรมความเข้าใจมากที่สุด ไม่ปรากฏข้อสอบที่วัดพฤติกรรมสังเคราะห์และประเมินผล

เนื้อหาการประเมินจิตพิสัย วัดพฤติกรรมความรู้มากที่สุด ไม่ปรากฏข้อสอบที่วัดพฤติกรรมความเข้าใจและการสังเคราะห์

เนื้อหาการประเมินทักษะการปฏิบัติ วัดพฤติกรรมความรู้มากที่สุด ไม่ปรากฏข้อสอบที่วัดพฤติกรรมการประเมิน

เนื้อหาการตัดสินผลการเรียนรู้ วัดพฤติกรรมวิเคราะห์มากที่สุด ไม่ปรากฏข้อสอบที่วัดพฤติกรรมการนำไปใช้ การสังเคราะห์ และการประเมิน

เนื้อหาการประเมินคุณภาพภายในและภายนอก วัดพฤติกรรมความรู้มากที่สุด ไม่ปรากฏข้อสอบที่วัดพฤติกรรมการนำไปใช้ การสังเคราะห์และการประเมิน

จุฬาลงกรณ์มหาวิทยาลัย

ตอนที่ 3 ผลการวิเคราะห์ค่าความเที่ยงของแบบสอบ

ผลการวิเคราะห์แบบสอบทั้งฉบับ จำแนกตามปีการศึกษาและภาคเรียน จำนวน 6 ฉบับมีจำนวนแบบสอบเลือกตอบ 65 ข้อ ส่วนใหญ่มีค่าเฉลี่ยคะแนนสอบใกล้เคียงกัน โดยปีการศึกษา 2546 ปลายภาค มีค่าเฉลี่ยคะแนนสูงที่สุด 43.77 รองลงมาเป็น กลางภาคเรียน 42.50 โดยปีการศึกษา 2544 ปลายภาคมีค่าเฉลี่ยต่ำมากที่สุด 38.55 ค่าส่วนเบี่ยงเบนมาตรฐานของข้อสอบทั้ง 6 ฉบับ มีค่าใกล้เคียงกันอยู่ในช่วงคะแนน 6.59-7.66 ค่าความเที่ยงแบบ KR-20 ของแบบสอบปีการศึกษา 2546 ปลายภาคมีค่าสูงที่สุด เท่ากับ 0.81 โดยแบบสอบปีการศึกษา 2544 ปลายภาคมีค่าต่ำที่สุดเท่ากับ 0.72 ค่าความคลาดเคลื่อนมาตรฐานของแบบสอบทั้งหมดมีความใกล้เคียงกันอยู่ในช่วง 3.30-3.54

ตารางที่ 10 ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ความเที่ยงแบบ KR-20 และความคลาดเคลื่อนมาตรฐาน จำแนกตามปีการศึกษาและภาคการเรียน

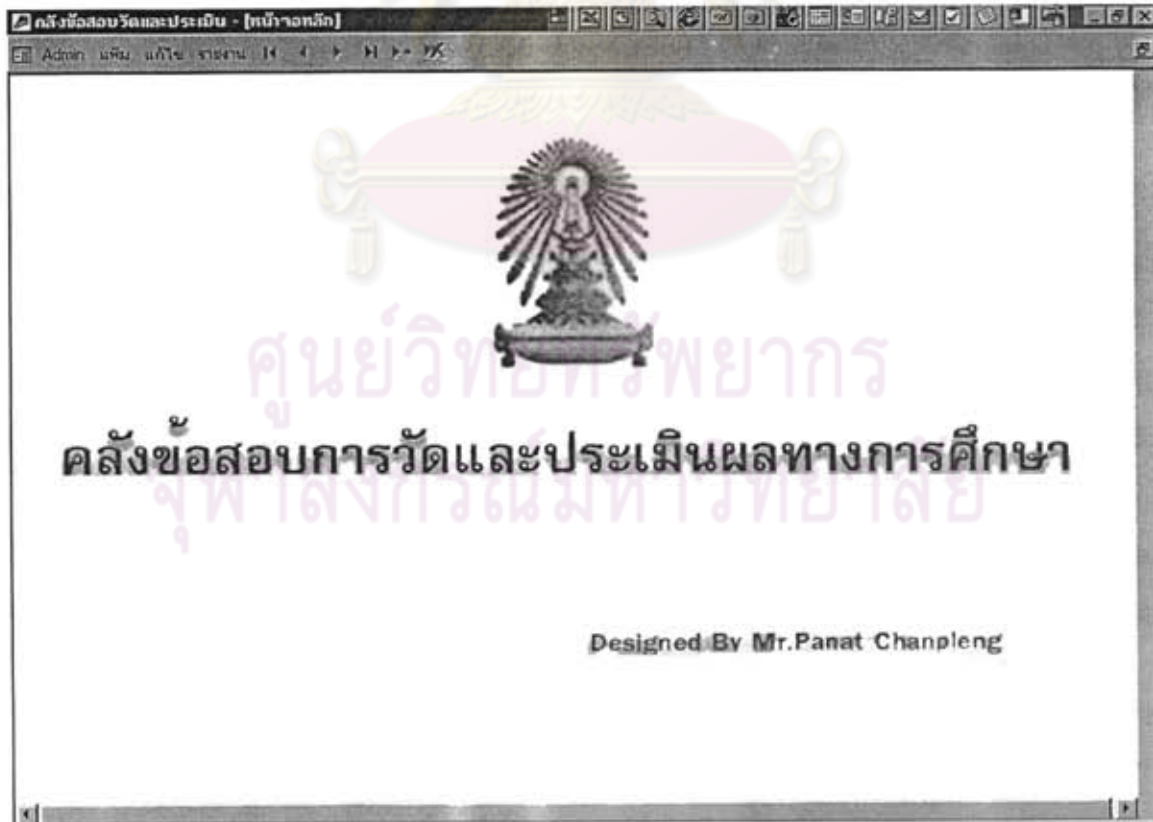
ปีการศึกษาและภาคการเรียน	จำนวนข้อ	Mean	S.D.	ความเที่ยง	SEM
1. พ.ศ. 2544 กลางภาค	65	40.41	6.70	0.77	3.30
2. พ.ศ. 2544 ปลายภาค	65	38.55	6.59	0.72	3.48
3. พ.ศ. 2545 กลางภาค	65	38.66	6.75	0.73	3.54
4. พ.ศ. 2545 ปลายภาค	65	41.96	7.10	0.77	3.37
5. พ.ศ. 2546 กลางภาค	65	42.50	6.76	0.74	3.46
6. พ.ศ. 2546 ปลายภาค	65	43.77	7.66	0.81	3.31

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตอนที่ 4 ลักษณะคลังข้อสอบวิชา 2702303 การวัดและการประเมินผลทางการศึกษา

โปรแกรมคอมพิวเตอร์ที่ใช้พัฒนาค้างข้อสอบวิชา 2702303 การวัดประเมินผลการศึกษา คือ โปรแกรมไมโครซอฟท์แอคเซส (Microsoft access) โดยผู้วิจัย และพนัส จันทรเปล่ง ได้ร่วมกัน วางรูปแบบและพัฒนาโปรแกรมคลังข้อสอบนี้ขึ้น เพื่อเก็บรวบรวมข้อสอบที่มีคุณภาพสำหรับใช้ในการวัดประเมินผลทางการศึกษา วิธีดำเนินการและผลการพัฒนาค้างข้อสอบมี 9 ขั้นตอน ได้แก่

1. วางแผนและกำหนดลักษณะคลังข้อสอบวิชา 2702303 การวัดและการประเมินผลการศึกษา ประกอบด้วยข้อมูลเกี่ยวกับข้อสอบ ดังนี้ 1) ลำดับของข้อคำถาม 2) ปีการศึกษา 3) ภาคการศึกษา 4) ข้อคำถาม/โจทย์ 5) คำตอบหรือตัวเลือก 6) ค่าความยาก 7) คำอธิบายแปลผลค่าความยาก 8) คำอำนาจจำแนก 9) คำอธิบายแปลผลอำนาจจำแนก 10) พฤติกรรมการเรียนรู้ 11) เนื้อหารายวิชา
2. จำแนกข้อสอบตามลักษณะเฉพาะดังกล่าวข้างต้นโดยผู้สอนประจำรายวิชา
3. คัดเลือกข้อสอบที่มี ค่าความยากอยู่ในช่วง 0.2-0.8 และคำอำนาจจำแนกตั้งแต่ 0.2 ขึ้นไป เพื่อจัดเก็บในโปรแกรมคลังข้อสอบ
4. ออกแบบโปรแกรมคอมพิวเตอร์คลังข้อสอบโดยเลือกให้โปรแกรม Microsoft access
5. เขียนโปรแกรม Microsoft access ซึ่งมีหน้าจอดังนี้



ภาพที่ 2 หน้าจอโปรแกรมคลังข้อสอบ

6. การจัดเก็บข้อสอบเข้าคลังข้อสอบ โดยคลิกเมาส์ที่หัวข้อ **เพิ่ม** และคลิก **กรอกข้อมูล** จะปรากฏหน้าจอเพื่อจัดเก็บข้อสอบที่คัดเลือกแล้วว่ามีความเหมาะสม ลงในโปรแกรม ดังนี้

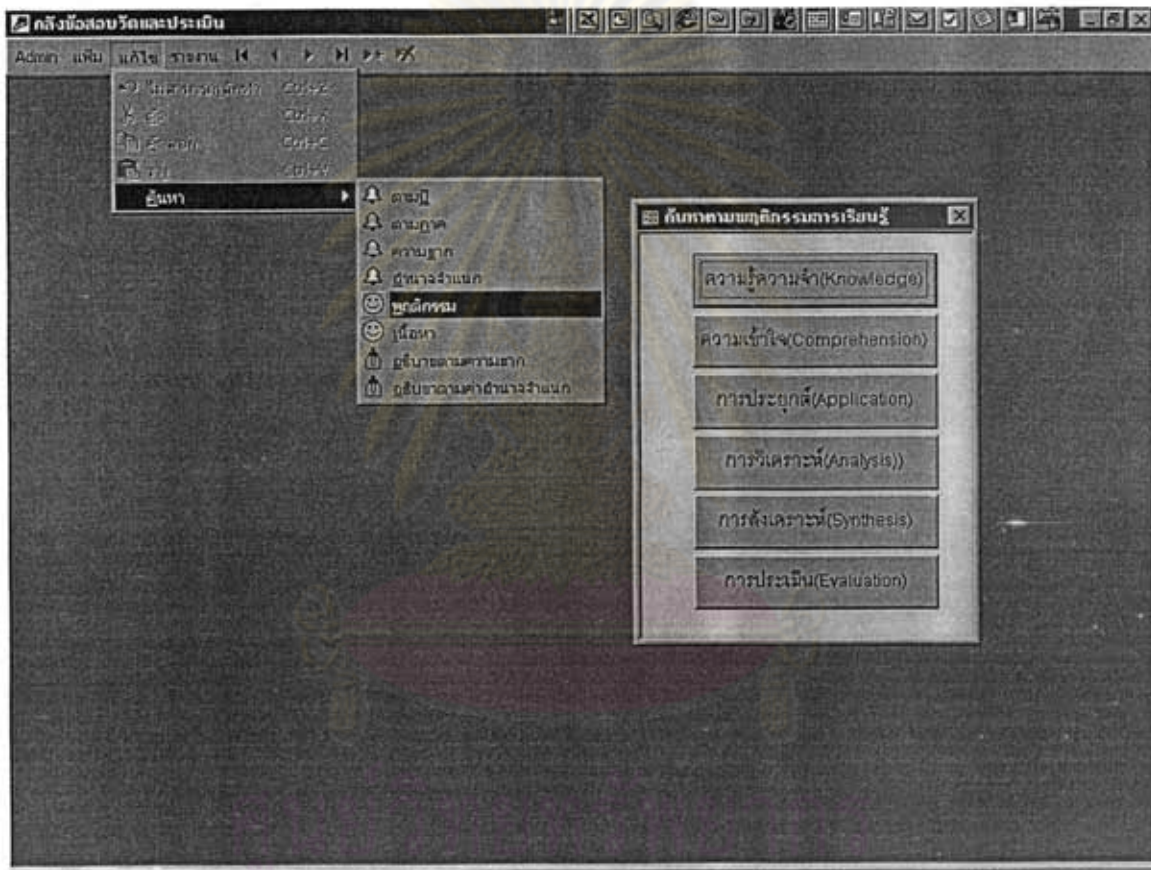
ภาพที่ 3 การจัดเก็บข้อสอบ

ในกรณีที่ผู้กรอกข้อมูลมีข้อสอบใหม่ ให้คลิก **ระเบียบ** มุมซ้ายมือด้านล่างเพื่อกำหนด ลำดับตัวเลขข้อที่จะเพิ่มเข้ามาใหม่ ช่อง ▶ * แล้วเริ่มกรอกข้อสอบใหม่บรรจุในคลังข้อสอบ การกรอกข้อมูลลงในช่องว่างที่กำหนดให้ตอบ เช่น ลำดับข้อ ปีการศึกษา ภาคการศึกษา โจทย์ ตัวเลือก ค่าความยาก ค่าอำนาจจำแนก พฤติกรรมการเรียนรู้ และเนื้อหา เป็นต้น แล้วดำเนินการ บันทึกข้อมูล

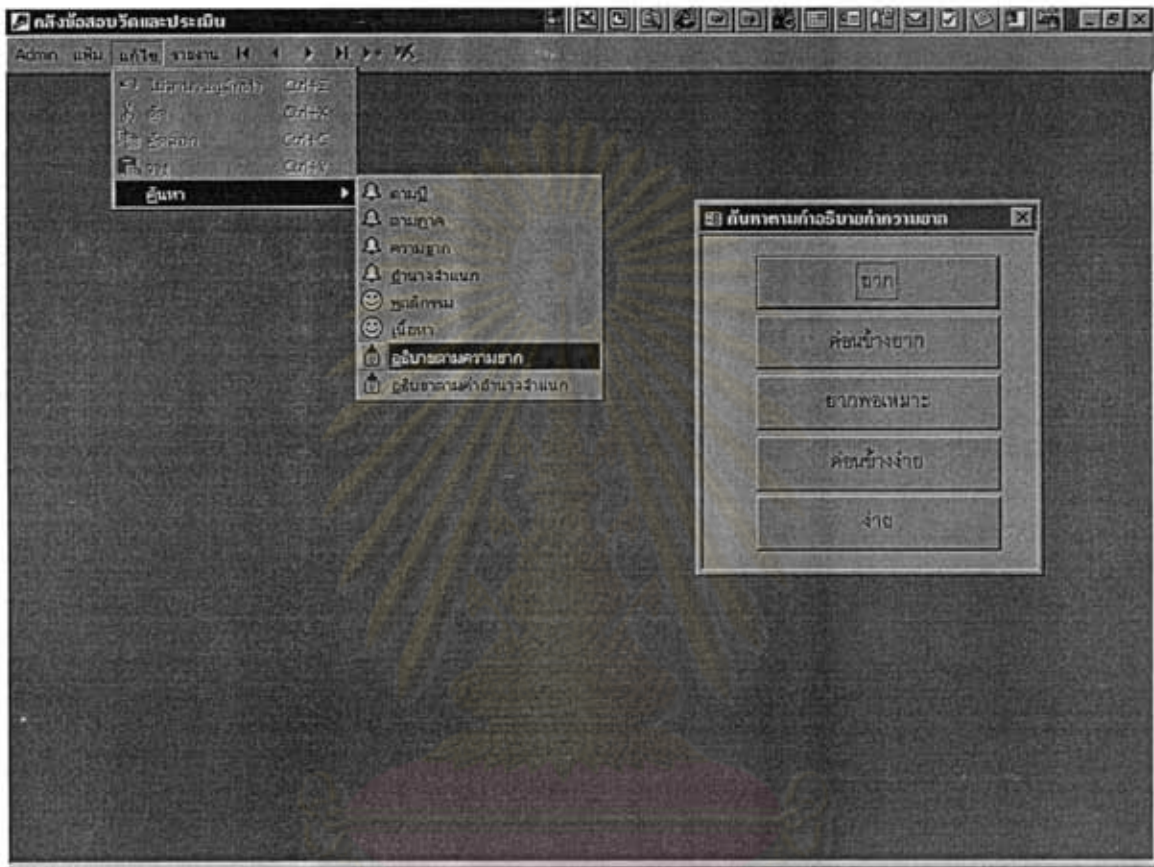
ในหัวข้อ **เพิ่ม** จากหน้าจอ โปรแกรมได้จำแนกข้อสอบตามค่าความยากและค่าอำนาจจำแนก แต่ละประเภทเป็น 5 กลุ่ม เพื่อเปิดโอกาสให้มีการจัดเก็บข้อสอบที่มีลักษณะที่แตกต่างจากเกณฑ์ที่กำหนดไว้ในอนาคต

7. คัดเลือกข้อสอบมาใช้งาน โดยกำหนดเลือกข้อสอบจาก ปีการศึกษา ภาคการศึกษา ค่าความยาก ค่าอำนาจจำแนก พฤติกรรมการเรียนรู้ และเนื้อหาวิชา

โดยการคลิก แก้ไข เลือกหัวข้อ ค้นหา แล้วคลิกเลือกตามลักษณะข้อสอบที่ต้องการ ได้แก่ ปีการศึกษา ภาคการศึกษา ค่าความยาก ค่าอำนาจจำแนก พฤติกรรมการเรียนรู้ และเนื้อหา มีคลิกเลือกหัวข้อที่ต้องการค้นหาแล้วจะปรากฏกล่องรายละเอียดที่ต้องการค้นหา ผู้ใช้งานสามารถ คลิก เลือกลักษณะข้อสอบตามรายการที่กำหนด ดังตัวอย่างภาพที่ 4-7 จากนั้นคลิก ข้อสอบจะแสดงจอภาพผลลัพธ์ของการค้นหาในภาพที่ 8

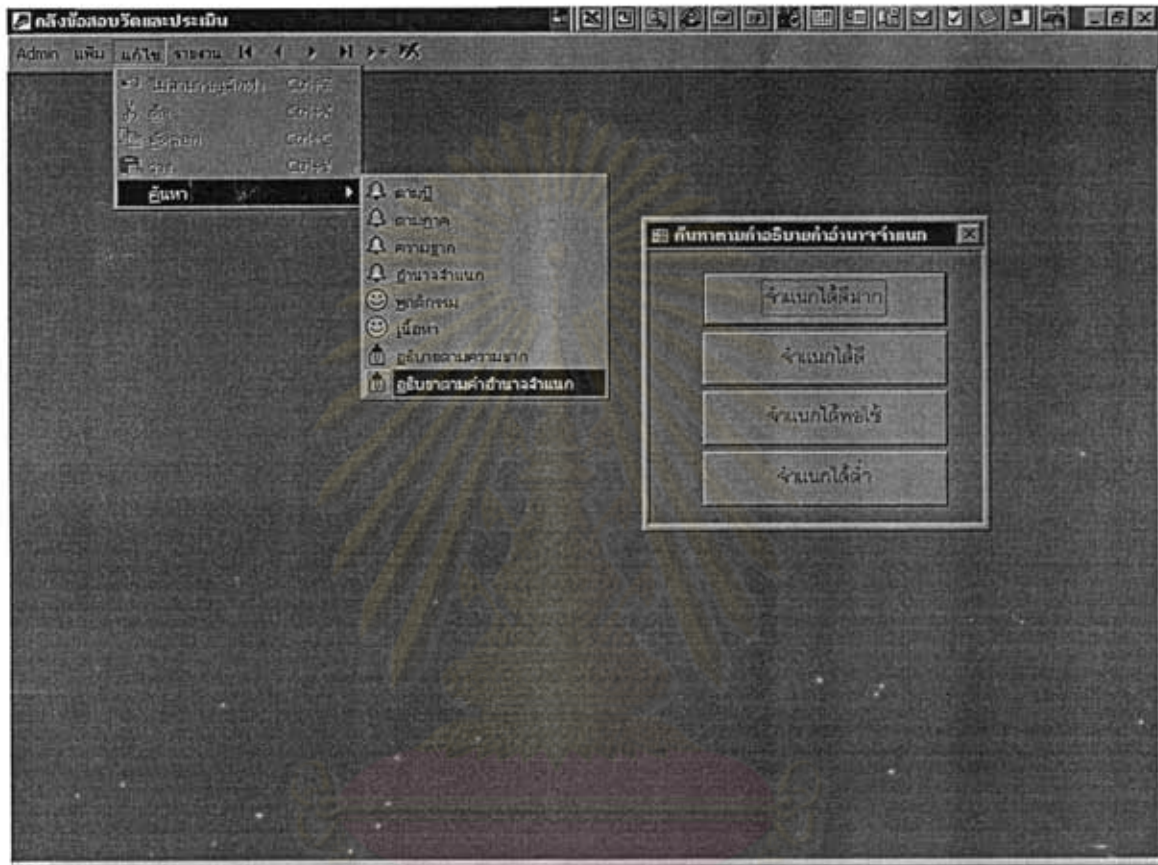


จุฬาลงกรณ์มหาวิทยาลัย
ภาพที่ 4 การค้นหาตามพฤติกรรมการเรียนรู้



ภาพที่ 6 การค้นหาตามความยาก

ศูนย์วิจัยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย



ภาพที่ 7 การค้นหาตามอำนาจจำแนก

ศูนย์วิจัยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ภาพที่ 8 ข้อมูลที่ได้จากการค้นหา

8. การจัดพิมพ์ข้อสอบโดยมีให้เลือกจัดพิมพ์ โดยการคลิกไปที่ รายงาน ข้อมูลการจัดพิมพ์มี 2 ลักษณะ ได้แก่

8.1 จัดพิมพ์ให้ปรากฏข้อคำถาม ตัวเลือก ค่าความยาก คำอธิบายแปลผล ค่าความยาก คำอำนาจจำแนก คำอธิบายแปลผลค่าอำนาจจำแนก พฤติกรรมการเรียนรู้ และเนื้อหาวิชา ในโปรแกรมคลังข้อสอบที่พัฒนาขึ้น

8.2 จัดพิมพ์ในโปรแกรม Microsoft word ซึ่งสามารถปรับแก้ เพิ่มเติม หรือ ลบ ข้อมูลได้ในแฟ้มข้อมูลดังกล่าว


คลังข้อสอบวัดและประเมินผล - [ตัวอย่างก่อนพิมพ์]

Admin เพิ่ม แก้ไข รายงาน

เลือกพื้นที่

ฐานงานทั้งหมด

ฐานค่าเฉลี่ย



ข้อสอบวิชาการวัดและประเมินผลทางการศึกษา

ข้อ 1 จุดเน้นของการวัดคืออะไร

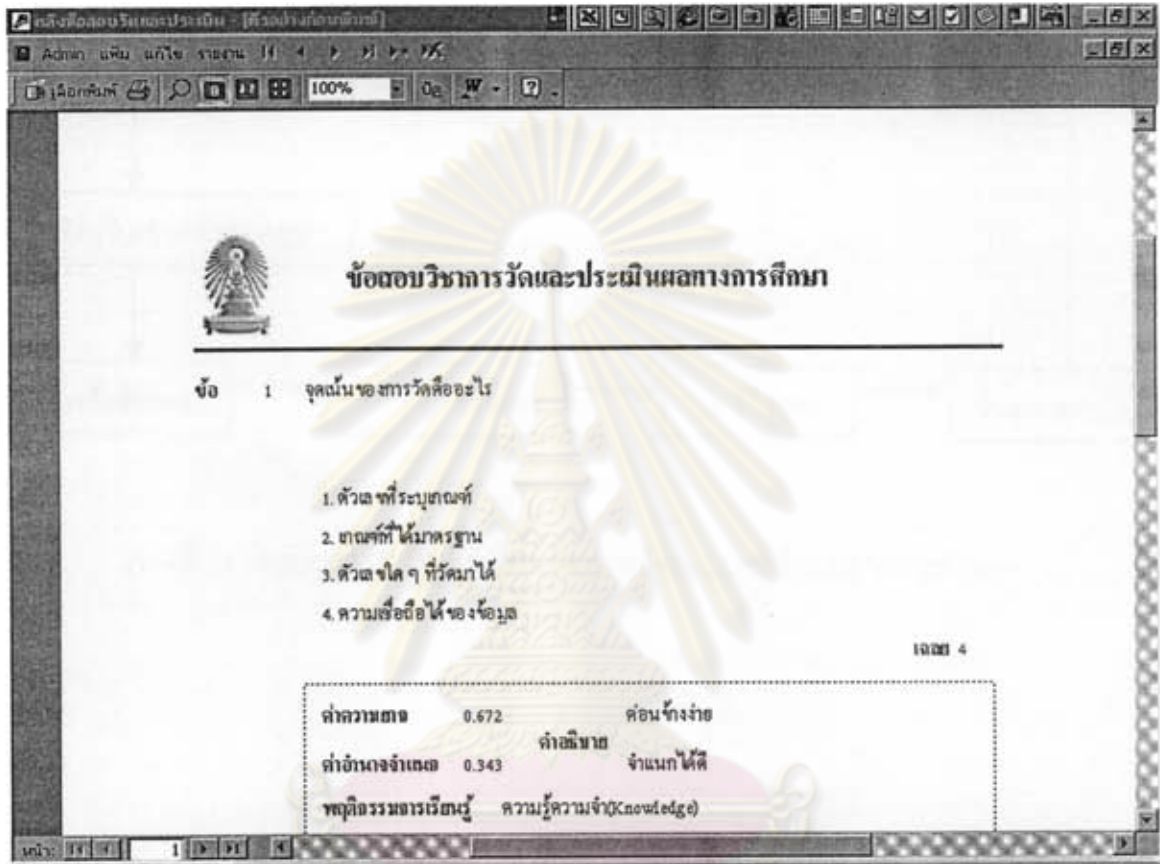
- ตัวเลขที่ระบุเกณฑ์
- ขาดสิ่งที่ได้มาตรฐาน
- ตัวเลขใด ๆ ที่วัดกันได้
- ความเชื่อถือได้ของข้อมูล

เฉลย 4

ค่าความยาก	0.672	ข้อนี้ง่าย
ค่าอำนาจจำแนก	0.343	จำแนกได้ดี
พฤติกรรมเฉลียว	ความรู้ความเข้าใจ(Knowledge)	

ภาพที่ 9 การจัดพิมพ์ในโปรแกรมคลังข้อสอบ

หากผู้ใช้ต้องการให้ข้อสอบที่ค้นหาจัดเก็บเพิ่มข้อมูลในรูปแบบโปรแกรม Microsoft word ให้คลิกที่ตัวสัญลักษณ์ W (Icon W) เพื่อเชื่อมโยงเพิ่มข้อมูล



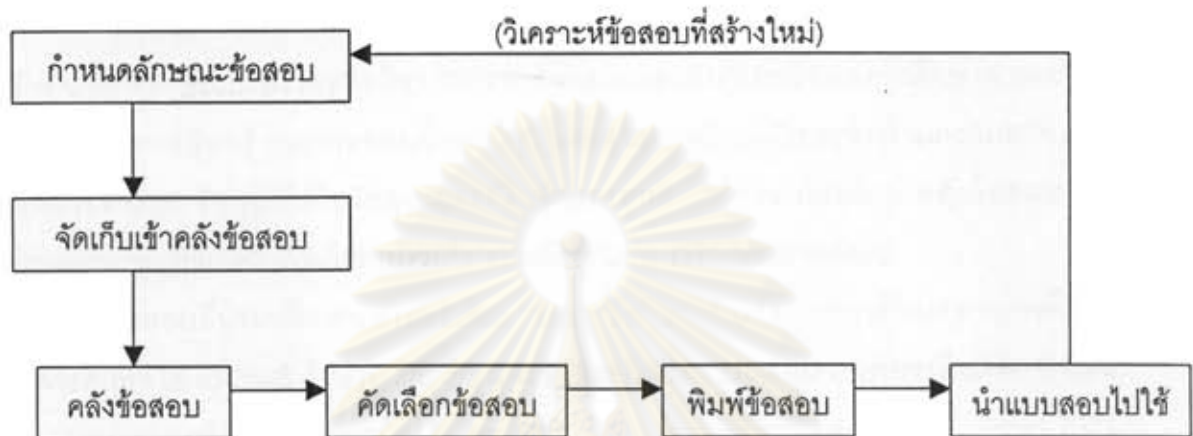
ภาพที่ 10 แฟ้มข้อมูลที่จัดเก็บในโปรแกรม Microsoft word

9. การรักษาความปลอดภัยของข้อสอบโดยกำหนดรหัส Password ในการเปิดโปรแกรม

ใช้งาน

จุฬาลงกรณ์มหาวิทยาลัย

กล่าวโดยสรุปคลังข้อสอบวิชา 2702303 การวัดและประเมินผลทางการศึกษามีขั้นตอนการใช้งานสำคัญๆ จะเกี่ยวข้องกับ การจัดเก็บข้อสอบเข้าคลังข้อสอบ การคัดเลือกข้อสอบมาใช้งานตามลักษณะของข้อสอบ และการจัดพิมพ์ข้อสอบเพื่อใช้งาน ดังภาพที่ 11



ภาพที่ 11 กระบวนการใช้คลังข้อสอบวิชาการวัดและประเมินผลทางการศึกษา

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 5

สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

การวิจัยครั้งนี้ใช้ระเบียบวิธีวิจัยเชิงบรรยาย มีวัตถุประสงค์สำคัญ 2 ข้อ คือ เพื่อศึกษาและจำแนกคุณลักษณะของข้อสอบวิชา 2702303 การวัดและการประเมินผลการศึกษาตามเกณฑ์พฤติกรรมการเรียนรู้ ประเภทของเนื้อหาวิชา และคุณภาพข้อสอบรายข้อด้านความยาก และอำนาจจำแนก ข้อมูลที่ได้ในวัตถุประสงค์ในข้อแรกถูกนำไปใช้ในเพื่อพัฒนาค้างข้อสอบ โดยโปรแกรมคอมพิวเตอร์ เพื่อเก็บรวบรวมข้อสอบที่มีคุณภาพไว้ใช้ในโอกาสต่อไป

ข้อสอบที่นำมาศึกษาประกอบด้วย ข้อสอบวิชา 2702303 การประเมินผลทางการศึกษา ปีการศึกษา 2545-2546 จำนวน 260 ข้อ และข้อสอบในวิชา 2702302 การประเมินผลการเรียนการสอน เฉพาะที่มีเนื้อหาเกี่ยวข้องกับวิชา 2702303 ประกอบด้วย ข้อสอบในปีการศึกษา 2541 จำนวน 207 และปีการศึกษา 2544 จำนวน 130 ข้อ ผู้วิจัยคัดเลือกข้อสอบทั้งหมดมาศึกษาจำนวน 597 ข้อ ข้อสอบทั้งหมดผ่านการวิเคราะห์ค่าความยากและอำนาจจำแนกแล้ว แล้วนำมาวิเคราะห์ข้อมูลจำแนกคุณลักษณะของข้อสอบ ตามเกณฑ์พฤติกรรมการเรียนรู้ ประเภทของเนื้อหา ด้วยสถิติความถี่ ร้อยละ และวิเคราะห์คุณภาพข้อสอบรายข้อด้านความยากและอำนาจจำแนก ด้วยการคำนวณค่าเฉลี่ยเลขคณิต ส่วนเบี่ยงเบนมาตรฐาน ค่าความเบ้ ค่าความโด่ง เป็นต้น

สรุปผลการวิจัย

ตอนที่ 1 ผลการวิเคราะห์ข้อสอบ

1.1 ข้อสอบวิชา 2702303 การวัดและการประเมินผลทางการศึกษา จำนวนทั้งสิ้น 597 ข้อ เป็นการวัดพฤติกรรมการเรียนรู้ในด้านความเข้าใจมากที่สุด ร้อยละ 31 รองลงมาเป็นความรู้ความจำ ร้อยละ 28.1 การวิเคราะห์ ร้อยละ 26 การนำไปใช้ ร้อยละ 10 การสังเคราะห์ ร้อยละ 3 สำหรับการวัดพฤติกรรมการเรียนรู้ด้านการประเมินมีจำนวนน้อยที่สุด ร้อยละ 1.8

1.2 เนื้อหารายวิชาที่มีจำนวนข้อสอบมากที่สุดคือ หลักการวัดและการประเมินผล/การวัดประเมินผลตามพระราชบัญญัติการศึกษา 2542 ร้อยละ 15.9 รองลงมาคือ การตรวจสอบคุณภาพเครื่องมือทั้งฉบับ ร้อยละ 13.7 และ การตรวจสอบคุณภาพเครื่องมือรายข้อ ร้อยละ 12.7 เนื้อหาที่มีข้อสอบวัดน้อยที่สุดคือ ทฤษฎี IRT และการสร้างธนาคารข้อสอบ ร้อยละ 3.9

1.3 ข้อสอบมีค่าความยากเฉลี่ย เท่ากับ 0.63 ซึ่งถือว่าข้อสอบค่อนข้างง่าย ส่วนเบี่ยงเบนมาตรฐานมีค่าเท่ากับ 0.21 มีลักษณะเบ้ซ้าย คือข้อสอบส่วนใหญ่ค่อนข้างง่าย ความโด่งมีลักษณะแบน คุณภาพความยากของข้อสอบทั้งหมดส่วนใหญ่เป็นข้อสอบค่อนข้างง่าย มีจำนวนมากที่สุด ร้อยละ 36.9 รองลงมาคือ ข้อสอบยากพอเหมาะ ร้อยละ 25.6 ข้อสอบที่สามารถเก็บไว้ใช้ได้ต่อไป คือ ข้อสอบค่อนข้างง่าย ยากพอเหมาะ และค่อนข้างยาก ซึ่งมีค่าความยากอยู่ในช่วง 0.2 – 0.8 มีจำนวนทั้งสิ้น 446 ข้อ คิดเป็นร้อยละ 74.7

1.4 ค่าอำนาจจำแนกเฉลี่ย เท่ากับ 0.26 ซึ่งเป็นค่าอำนาจจำแนกได้พอใช้ ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.16 มีลักษณะเบ้ซ้าย และความโด่งมีลักษณะแบน คุณภาพด้านอำนาจจำแนกของข้อสอบทั้งหมดส่วนใหญ่เป็นข้อสอบจำแนกได้ต่ำมีจำนวนมากที่สุด ร้อยละ 31.5 รองลงมาคือ ข้อสอบจำแนกพอใช้ ร้อยละ 23.5 ข้อสอบที่สามารถเก็บไว้ใช้ได้ต่อไป คือ ข้อสอบที่มีอำนาจจำแนกได้พอใช้ จำแนกได้ดี และจำแนกได้ดีมา ซึ่งมีค่าอำนาจจำแนกตั้งแต่ 0.2 เป็นต้นไป จำนวนทั้งสิ้น 386 ข้อ คิดเป็นร้อยละ 64.5

1.5 ข้อสอบที่ผ่านเกณฑ์ความยากซึ่งมีค่าอยู่ในช่วง 0.2 – 0.8 และค่าอำนาจจำแนกตั้งแต่ 0.2 ขึ้นไป ถือเป็นข้อสอบที่มีคุณภาพและเก็บไว้ใช้ได้ต่อไป มีจำนวนทั้งสิ้น 322 ข้อ คิดเป็นร้อยละ 53.9

ตอนที่ 2 ผลการวิเคราะห์ข้อสอบที่จะคัดเลือกไว้ในคลังข้อสอบ

2.1 ข้อสอบที่เก็บไว้ใช้ เป็นข้อสอบที่วัดพฤติกรรมการเรียนรู้ในด้านความเข้าใจมากที่สุด ร้อยละ 32.3 รองลงมาเป็นความรู้ความจำ ร้อยละ 29.2 การวิเคราะห์ ร้อยละ 26.1 การนำไปใช้ ร้อยละ 7.1 การสังเคราะห์ ร้อยละ 3.4 สำหรับการวัดพฤติกรรมการเรียนรู้ด้านการประเมินมีจำนวนน้อยที่สุด ร้อยละ 1.9

2.2 เนื้อหารายวิชาที่มีจำนวนข้อสอบที่มีคุณภาพใช้ได้มากที่สุดคือ การวิเคราะห์คุณภาพแบบสอบทั้งฉบับ ร้อยละ 19.6 รองลงมาคือ การวิเคราะห์คุณภาพรายข้อ ร้อยละ 14.9 และหลักการวัดและการประเมินผล/การวัดประเมินผลตามพระราชบัญญัติการศึกษา 2542 ร้อยละ 13.7 เนื้อหาที่มีข้อสอบที่มีคุณภาพจำนวนน้อยที่สุดคือ การประเมินทักษะการปฏิบัติ ร้อยละ 3.1

ตอนที่ 3 ผลการวิเคราะห์ค่าความเที่ยงของแบบสอบ

ค่าความเที่ยงแบบ KR-20 ของแบบสอบปีการศึกษา 2546 ปลายภาคมีค่าสูงที่สุด เท่ากับ 0.81 โดยแบบสอบปีการศึกษา 2544 ปลายภาคมีค่าต่ำที่สุดเท่ากับ 0.72 ค่าความคลาดเคลื่อนมาตรฐานของแบบสอบทั้งหมดมีความใกล้เคียงกัน อยู่ในช่วง 3.30-3.54

ตอนที่ 4 ลักษณะคลังข้อสอบวิชา 2702303 การวัดและการประเมินผลทางการศึกษา

โปรแกรมคอมพิวเตอร์ที่ใช้พัฒนาค้นข้อสอบวิชา 2702303 การวัดประเมินผลการศึกษา คือ โปรแกรมไมโครซอฟต์แอสเซส (Microsoft access) โดยผู้วิจัย และพนัส จันทร์เปล่ง ได้ร่วมกัน วางรูปแบบและพัฒนาโปรแกรมคลังข้อสอบนี้ขึ้น เพื่อเก็บรวบรวมข้อสอบที่มีคุณภาพสำหรับใช้ในการ วัดประเมินผลทางการศึกษา วิธีดำเนินการและผลการพัฒนาค้นข้อสอบมี 9 ขั้นตอน ได้แก่

1. การวางแผนและกำหนดลักษณะคลังข้อสอบวิชา 2702303 การวัดและการประเมินผลทางการศึกษา ประกอบด้วยข้อสอบที่มีข้อมูลเกี่ยวกับ 1) ลำดับของข้อคำถาม 2) ปีการศึกษา 3) ภาคเรียน 4) ข้อคำถาม/โจทย์ 5) คำตอบหรือตัวเลือก 5 ตัวเลือก 6) ค่าความยาก 7) คำอธิบายแปลผลค่าความยาก 8) คำอำนาจจำแนก 9) คำอธิบายแปลผลอำนาจจำแนก 10) พฤติกรรม การเรียนรู้ 11) เนื้อหาวิชา
2. การจำแนกข้อสอบตามลักษณะตัวแปรดังกล่าวข้างต้น
3. การคัดเลือกข้อสอบที่มีค่าความยากอยู่ในช่วง 0.2-0.8 และคำอำนาจจำแนกตั้งแต่ 0.2 ขึ้นไป เพื่อจัดเก็บในโปรแกรม
4. การออกแบบโปรแกรมคอมพิวเตอร์ระบบปฏิบัติการ Windows โดยเลือกใช้โปรแกรม Microsoft access
5. การเขียนคำสั่งโปรแกรม Microsoft access
6. การจัดเก็บข้อสอบเข้าคลังข้อสอบ
7. การคัดเลือกข้อสอบมาใช้งาน โดยกำหนดเลือกข้อสอบจาก ปีการศึกษา ภาคเรียน ค่าความยาก อำนาจจำแนก พฤติกรรมการเรียนรู้ และเนื้อหาวิชา
8. การจัดพิมพ์แบบสอบโดยมีให้เลือกจัดพิมพ์ 2 ลักษณะคือ
 - 8.1 จัดพิมพ์ให้ปรากฏข้อคำถาม ตัวเลือก ค่าความยาก คำอธิบายแปลผลค่าความยาก คำอำนาจจำแนก คำอธิบายแปลผลอำนาจจำแนก พฤติกรรมการเรียนรู้ และ เนื้อหาวิชา
 - 8.2 จัดพิมพ์ในโปรแกรม Microsoft word
9. การรักษาความปลอดภัยของข้อสอบโดยกำหนดรหัส Password ในการเปิดโปรแกรมใช้งาน

อภิปรายผล

ผลการพัฒนาคลังข้อสอบวิชา 2702303 การวัดและการประเมินผลทางการศึกษาด้วยโปรแกรมคอมพิวเตอร์ มีประเด็นน่าสนใจต่อการอภิปรายผล ดังนี้

1. การจำแนกคุณลักษณะของข้อสอบตามพฤติกรรมการเรียนรู้

ผลที่ได้จากการศึกษาพบว่า ข้อสอบจำนวน 597 ข้อที่นำมาศึกษา ส่วนใหญ่วัดพฤติกรรมการเรียนรู้ ด้านความเข้าใจมากที่สุด ร้อยละ 31 รองลงมาเป็นเรื่องความรู้ความจำ การวิเคราะห์ การนำไปใช้ การสังเคราะห์ และการประเมินค่า ตามลำดับ ซึ่งสอดคล้องกับจุดมุ่งหมายของวิชา เมื่อนิสิตได้ศึกษาครบทุกเนื้อหาแล้ว สามารถอธิบายหลักการวัดและประเมินผลการเรียนรู้ได้ถูกต้อง วิเคราะห์ข้อสอบและแบบสอบได้ สร้างและพัฒนาแบบสอบผลสัมฤทธิ์ที่เชื่อถือได้ ตรวจสอบให้คะแนนและอันดับคะแนนได้ พร้อมทั้งประยุกต์หลักการประเมินผลมาใช้ในการเรียนการสอนได้ อย่างไรก็ตามผลการวิจัยครั้งนี้ใช้เป็นข้อมูลให้ผู้สอนได้พิจารณาสร้างข้อสอบที่มีระดับการวัดพฤติกรรมการเรียนรู้ระดับสูงขึ้น สำหรับการสร้างแบบสอบเพื่อวัดและประเมินผลในอนาคต

2. การจำแนกคุณลักษณะของข้อสอบตามเนื้อหาวิชา

ผลที่ได้จากการศึกษาพบว่า เนื้อหาที่มีข้อสอบจำนวนมากที่สุดคือ หลักการวัดและการประเมินผล/ การวัดและการประเมินผลตามพระราชบัญญัติการศึกษาแห่งชาติ 2542 (15.9%) รองลงมา คือ การตรวจสอบคุณภาพเครื่องมือทั้งฉบับ (13.7%) การตรวจสอบคุณภาพเครื่องมือรายข้อ (12.7%) เป็นต้น ซึ่งสอดคล้องกับเนื้อหาตามประมวลรายวิชา 2702303 ซึ่งให้น้ำหนักความสำคัญกับเนื้อหาดังกล่าวมากกว่าด้านอื่น โดยเนื้อหาข้างต้นผู้สอนจะใช้เวลาในการสอนและออกข้อสอบจำนวนมากกว่าเนื้อหาอื่นๆ จากการกำหนดจำนวนข้อสอบในแผนผังข้อสอบของแบบสอบกลางภาคและปลายภาคจำนวน 130 ข้อ ได้ให้น้ำหนักกับเนื้อหาดังกล่าว 15 ข้อ คิดเป็นร้อยละ 11.54 และเนื้อหาอื่นๆ 10 ข้อ คิดเป็นร้อยละ 7.69

อย่างไรก็ตามเมื่อเปรียบเทียบระหว่างแผนผังการออกข้อสอบกับรายวิชา 2702303 ในปัจจุบัน พบว่า จำนวนตัวเลขไม่ตรงกับแผนผังข้อสอบในปัจจุบันอย่างแท้จริง ทั้งนี้เนื่องจากข้อสอบที่นำมาวิเคราะห์เป็นข้อที่นำมาจากรายวิชา 2702302 ปีการศึกษา 2541, 2544 และรายวิชา 2702303 ปีการศึกษา 2545-2546 เป็นข้อสอบที่มีแผนผังข้อสอบ (Testing blueprint) ที่กำหนดน้ำหนักข้อสอบ จำนวนข้อสอบแต่ละเนื้อหาแตกต่างกัน แต่ผู้วิจัยได้คัดเลือกมาเฉพาะข้อสอบที่มีเนื้อหาเกี่ยวข้องกับรายวิชา 2702303 เท่านั้น เพื่อคัดเลือกข้อสอบที่มีคุณภาพจัดเก็บในคลังข้อสอบ หากเปรียบเทียบจำนวนข้อสอบกับแผนผังข้อสอบที่ใช้ในปัจจุบันทำได้เฉพาะการจัดลำดับจำนวนมาน้อยของข้อสอบเท่านั้น

3. การจำแนกข้อสอบตามคุณภาพข้อสอบรายข้อ ด้านความยากและอำนาจจำแนก

ผลที่ได้จากการศึกษาพบว่า ข้อสอบจำนวน 597 ข้อ เมื่อวิเคราะห์คุณภาพข้อสอบรายข้อ ด้วยการพิจารณาค่าความยากและค่าอำนาจจำแนกตามแนวคิดการประเมินแบบอิงกลุ่ม ข้อสอบที่ผ่านเกณฑ์ความยากซึ่งมีค่าอยู่ในช่วง 0.2 – 0.8 และค่าอำนาจจำแนก ตั้งแต่ 0.2 ขึ้นไป ถือเป็นข้อสอบที่มีคุณภาพและเก็บไว้ใช้ได้ต่อไป มีจำนวนทั้งสิ้น 322 ข้อ คิดเป็นร้อยละ 53.94 ด้วยข้อจำกัดของโปรแกรมที่ไม่สามารถจัดเก็บข้อสอบที่มีลักษณะเป็นตาราง แผนภูมิ สัญลักษณ์ทางสถิติ และรูปภาพประกอบ จึงทำให้มีข้อสอบที่สามารถบันทึกในโปรแกรมคลังข้อสอบจำนวน 262 ข้อ คิดเป็นร้อยละ 43.89

4. ความสัมพันธ์ระหว่างค่าความยากและค่าอำนาจจำแนก

ผลการศึกษาการกระจายค่าความยากและค่าอำนาจจำแนกของข้อสอบ พบว่า ข้อสอบที่ง่ายมากหรือยากมากจะมีอำนาจจำแนกอยู่ในช่วงจำกัดและน้อย ข้อสอบที่มีอำนาจจำแนกตั้งแต่ 0.2 ขึ้นไป ส่วนใหญ่เป็นข้อสอบที่มีความยากค่อนข้างง่ายและปานกลาง จะมีค่าอำนาจจำแนกสูง ผลการวิจัยดังกล่าวมีความสอดคล้องกับคำอธิบายความสัมพันธ์ระหว่างความยากและค่าอำนาจจำแนกของข้อสอบของ Carey (1988; อ้างถึงใน ศิริชัย กาญจนวาสิ, 2544) ซึ่งเป็นข้อค้นพบนอกเหนือจากวัตถุประสงค์การวิจัยที่น่าสนใจและสอดคล้องตามหลักการวัดและประเมินผล

5. การพัฒนาคลังข้อสอบ ด้วยโปรแกรมคอมพิวเตอร์

การพัฒนาคลังข้อสอบในการวิจัยนี้มีขั้นตอนการพัฒนาคลังข้อสอบ 9 ขั้นตอน ตามกรอบแนวคิดที่ประยุกต์จาก ศิริชัย กาญจนวาสิ (2541) 7 ขั้นตอน ได้แก่ การวางแผน การออกแบบโปรแกรม การจัดเก็บข้อสอบเข้าคลัง การคัดเลือกข้อสอบจากคลังมาใช้งาน การจัดพิมพ์ข้อสอบ การให้สารสนเทศเกี่ยวกับข้อสอบ การรักษาความปลอดภัยของระบบคลังข้อสอบเป็นกรอบในการพัฒนาคลังข้อสอบ และด้วยคลังข้อสอบที่สร้างขึ้น ไม่สามารถตรวจให้คะแนนและวิเคราะห์ข้อสอบได้ จึงไม่มีขั้นตอนของการให้สารสนเทศเกี่ยวกับแบบสอบ อย่างไรก็ตามผู้วิจัยได้เพิ่มขั้นตอนการจำแนกข้อสอบตามลักษณะเฉพาะ การคัดเลือกข้อสอบ และการเขียนโปรแกรมเพิ่มเติม 3 ขั้นตอน รวมเป็น 9 ขั้นตอน ประกอบด้วย 1) การวางแผนกำหนดลักษณะคลังข้อสอบ 2) การจำแนกข้อสอบ 3) การคัดเลือกข้อสอบที่มีคุณภาพ 4) การออกแบบโปรแกรมคอมพิวเตอร์ 5) การเขียนคำสั่งโปรแกรมไมโครซอฟต์แอสแซส 6) การจัดเก็บข้อสอบในคลังข้อสอบ 7) การคัดเลือกข้อสอบมาใช้งาน 8) การจัดพิมพ์แบบสอบ และ 9) การรักษาความปลอดภัย

คลังข้อสอบที่พัฒนาขึ้นทำหน้าที่เพียงจัดเก็บข้อสอบที่ผ่านการวิเคราะห์คุณภาพข้อสอบรายข้อได้แก่ ค่าความยาก และค่าอำนาจจำแนก มาก่อน และการพิมพ์ข้อสอบที่เป็นแบบสอบ

เลือกตอบ จำนวนไม่เกิน 5 ตัวเลือก สำหรับสร้างแบบสอบกลางภาคและปลายภาคเรียนของรายวิชา 2702303 ซึ่งไม่ได้ดำเนินการวัดและการประเมินผลผ่านระบบปฏิบัติการทางคอมพิวเตอร์

การทำหน้าที่ของคลังข้อสอบเพียงการจัดเก็บและการพิมพ์ข้อสอบในลักษณะนี้เหมือนกับคลังข้อสอบจากการวิจัยที่ผ่านมา (ประคอง กรรณสูต, 2536; วินุลาศ เจริญชัย, 2537; วีรยุทธ ธานี, 2546) และแตกต่างจากงานวิจัยของ กฤษภาพันธ์ พงษ์บริบูรณ์ (2543) และ วินุลาศ เจริญชัย (2537) ซึ่งพัฒนาคลังข้อสอบที่สามารถจัดเก็บ และวิเคราะห์ข้อสอบได้ รวมทั้งงานวิจัยของ กอบกิจ สหัสรังษี (2539) ที่คลังข้อสอบทำหน้าที่จัดเก็บ ดำเนินการสอบ และวิเคราะห์ข้อสอบได้ในระบบปฏิบัติการคอมพิวเตอร์

ข้อเสนอแนะ

ผู้วิจัยแบ่งข้อเสนอแนะ ออกเป็น 2 หัวข้อย่อย คือ ข้อเสนอแนะในการนำผลวิจัยไปใช้ และ ข้อเสนอแนะในการวิจัยต่อไป โดยมีรายละเอียดแต่ละหัวข้อดังนี้

1. ข้อเสนอแนะในการนำผลวิจัยไปใช้

1. เนื่องจากคลังข้อสอบที่พัฒนาในการวิจัยนี้ ทำหน้าที่ 3 ประการสำคัญ คือ การจัดเก็บข้อสอบ การคัดเลือกข้อสอบมาใช้งานตามคุณสมบัติข้อสอบ และการจัดพิมพ์ข้อสอบไปใช้ แต่ไม่สามารถทำหน้าที่ด้านการวิเคราะห์ข้อสอบได้ ข้อสอบที่นำมาจัดเก็บจึงต้องผ่านการวิเคราะห์คุณภาพรายข้อด้าน ความยากและอำนาจจำแนก ตามแนวทฤษฎีการทดสอบแบบดั้งเดิมมาก่อนที่จะจัดเก็บลงในคลังข้อสอบนั้นได้

2 ข้อสอบที่จัดเก็บได้ดี เป็นข้อสอบแบบเลือกตอบเป็นลักษณะข้อความ ที่มีตัวเลือกไม่เกิน 5 เท่านั้น ไม่สามารถจัดเก็บข้อสอบเลือกตอบที่มีลักษณะคำถามต่อเนื่อง หรือคำถามที่มี ตารางสัญลักษณ์ทางสถิติ รูปภาพประกอบได้

3. เครื่องคอมพิวเตอร์ที่จะใช้คลังข้อสอบนี้ ต้องใช้ระบบปฏิบัติการ Windows 2000 ขึ้นไป

4. ผู้ใช้โปรแกรม ควรมีความรู้ความเข้าใจพื้นฐานเกี่ยวกับ เนื้อหาวิชาที่สอน ระดับพฤติกรรมที่วัด ความยาก อำนาจจำแนก และสามารถใช้คอมพิวเตอร์ได้เป็นอย่างดี

5. คณาจารย์ผู้สอนวิชา 2702303 การวัดและการประเมินผลทางการศึกษา ควรพัฒนาข้อสอบที่วัดพฤติกรรมการเรียนรู้ระดับ การนำไปใช้ การสังเคราะห์ และการประเมินผลเพิ่มขึ้น เนื่องจากมีข้อสอบที่สามารถวัดพฤติกรรมด้านนี้ในจำนวนน้อย เมื่อพิจารณาด้านเนื้อหาวิชา ครูผู้สอนควรเพิ่มเติมเกี่ยวกับ ทฤษฎี IRT และการสร้างธนาคารข้อสอบ การประเมินคุณภาพภายใน

และภายนอก กรประเมินทักษะการปฏิบัติ เพื่อให้มีข้อสอบจำนวนมากรองรับการนำไปใช้พัฒนาแบบสอบมากขึ้น

2. ข้อเสนอแนะในการวิจัยต่อไป

1. ในการวิจัยครั้งนี้ เป็นการวิจัยเชิงปฏิบัติการ พัฒนาคคลังข้อสอบขึ้นใช้จริงในการจัดการเรียนการสอนวิชาวัดและประเมินผลทางการศึกษา มีเป้าหมายสำคัญเพื่อการจัดเก็บข้อสอบที่มีคุณภาพในการใช้วัดประเมินผลในอนาคตเท่านั้น ซึ่งการทำหน้าที่ของคลังข้อสอบที่พัฒนาขึ้นไม่ได้ครอบคลุมถึงการวิเคราะห์ข้อสอบ เนื่องจากทางหน่วยงานในคณะโดยศูนย์ทดสอบคณะครุศาสตร์ ได้บริการวิเคราะห์ข้อสอบให้แล้ว การพัฒนาคคลังข้อสอบมีหน้าที่สมบูรณ์ต่อไปสำหรับผู้สนใจ คือ การเพิ่มความสามารถในการวิเคราะห์ข้อสอบในคลังข้อสอบที่ทำการศึกษาพัฒนาต่อไป

2. เนื่องจากข้อสอบเลือกตอบที่บรรจุได้ในคลังข้อสอบนี้ ต้องมีการตอบเป็นอิสระจากข้ออื่น ๆ และมีลักษณะคำถามเป็นข้อความเพียงอย่างเดียว ดังนั้นการทำวิจัยเพื่อพัฒนาโปรแกรมคลังข้อสอบนี้ควรพัฒนาขีดความสามารถในการบรรจุข้อสอบเลือกตอบที่มี ตาราง สัญลักษณ์ทางสถิติ รูปภาพประกอบได้ รวมทั้งความสามารถจัดเก็บข้อสอบเลือกตอบที่มีลักษณะข้อคำถามต่อเนื่องได้ต่อไป นอกจากนี้ตามหลักการวัดผลที่ดี ควรมีวิธีการและประเภทของข้อสอบที่ใช้วัดประเมินผลอย่างหลากหลาย คลังข้อสอบที่พัฒนาขึ้นก็ควรตอบสนองหลักการวัดผลดังกล่าว โดยเพิ่มหน้าที่บรรจุข้อสอบที่มีลักษณะต่างๆ เช่น จับคู่ ถูกผิด เติมคำ ตอบสั้น หรือ ความเรียงได้ในอนาคตต่อไป

3. การบรรจุคุณภาพข้อสอบรายข้อ ด้านความยาก และอำนาจจำแนก ด้วยหลักการวิเคราะห์ตามแนวทฤษฎีการทดสอบแบบดั้งเดิมมีจุดอ่อนที่ค่าพารามิเตอร์ของข้อสอบมีการเปลี่ยนแปลงไปตามกลุ่มผู้สอบ การพัฒนาคคลังข้อสอบที่เป็นมาตรฐานจะไม่ควรมีการแปรเปลี่ยนของค่าพารามิเตอร์ข้อสอบรายข้อ และการวิเคราะห์คุณภาพข้อสอบด้วยทฤษฎีการตอบสนองข้อสอบ จะช่วยลดปัญหาดังกล่าวลงได้ แต่มีข้อควรคำนึงว่าหากจะใช้หลักการวิเคราะห์ข้อสอบตามแนวทฤษฎีการตอบสนองข้อสอบ จะสามารถกระทำได้อย่างเหมาะสมในรายวิชาที่ผู้เรียนหรือผู้สอบมีจำนวนมาก

บรรณานุกรม

- กอบกิจ สหัสรังษี. (2539). *การพัฒนาระบบการจัดการข้อสอบปรนัย*. วิทยานิพนธ์วิทยาศาสตร์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ จุฬาลงกรณ์มหาวิทยาลัย.
- กฤษฎาพันธ์ พงษ์บริบูรณ์. (2543). *การพัฒนาโปรแกรมคอมพิวเตอร์ เพื่อการวิเคราะห์ข้อสอบ และการจัดเก็บ สำหรับแบบทดสอบอิงเกณฑ์*. วิทยานิพนธ์ศึกษาศาสตร์มหาวิทยาลัยขอนแก่น.
- คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย. (2545). *คู่มือนิสิตปริญญาบัณฑิต*. คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.
- ประคอง วรรณสุด. (2536). *รายงานการวิจัยเรื่องการพัฒนาคลังข้อสอบด้วยโปรแกรมประมวลผลคำโดยใช้ไมโครคอมพิวเตอร์*. ภาควิชาวิจัยการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.
- พลการ กรพิทักษ์. (2533). *การพัฒนาโปรแกรมคอมพิวเตอร์สำหรับการจัดเก็บข้อสอบโดยใช้ไมโครคอมพิวเตอร์*. วิทยานิพนธ์ครุศาสตร์มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ จุฬาลงกรณ์มหาวิทยาลัย.
- ภาวิณี ศรีสุขวัฒนานันท์. (2543). *ระบบการทดสอบด้วยคอมพิวเตอร์*. วารสารศึกษาศาสตร์ปริทัศน์. ปีที่ 15 ฉบับที่ 3. (ก.ย.-ธ.ค.) 65-71.
- วิบูลาศ เจริญชัย. (2537). *การพัฒนาโปรแกรมไมโครคอมพิวเตอร์จัดเก็บและวิเคราะห์ข้อสอบเลือกตอบ*. วิทยานิพนธ์ศึกษาศาสตร์มหาวิทยาลัยขอนแก่น.
- วีรยุทธ์ ธานี. (2546). *การพัฒนาโปรแกรมคลังข้อสอบบนระบบปฏิบัติการวินโดวส์*. วิทยานิพนธ์ครุศาสตร์มหาวิทยาลัย สถาบันราชภัฏอุบลราชธานี.
- ศิริชัย กาญจนวาสี. (2541). *เอกสารประกอบการสอนวิชา 2702639 ทฤษฎีการวัดและประเมิน*. คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย, (อัดสำเนา).
- ศิริชัย กาญจนวาสี. (2544). *ทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory)*. กรุงเทพมหานคร: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- สุพัฒน์ สุกมลลันต์. (2539). *ธนาคารข้อทดสอบการทดสอบ การทดสอบปรับเปลี่ยนด้วยคอมพิวเตอร์*. กรุงเทพมหานคร: บริษัทพิมพ์ดีจำกัด.
- Coppin, B. H. (1976). Recent developments in item banking; A review in D.N.M. DeGrujter and L.J.Th. van der Kamp (EDs.), *Advances in psychological and*

educational Measurement. London: Wiley & Sons.

Choppin, B. H. (1985). Item Banking Using Sample-Free Calibration. *Evaluation in Education An International Review Series*, Vol. 9.

Millman, J. and Arter, J. A. (1984). Issues in item banking. *Journal of Educational Measurement*. Vol 21, No. 4.

Wright, B. D., and Bell, S. R. (1984). Item bank: what, why, how. *Journal of Educational measurement*. Vol 21, No. 4.



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย



ภาคผนวก

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

รายชื่อผู้ดำเนินการวิจัย

ที่ปรึกษา รศ.ดร.ศิริชัย กาญจนวาสี

นักวิจัย อ.ดร.ณัฐภรณ์ หลาวทอง

ผู้ช่วยนักวิจัยทำหน้าที่เขียนโปรแกรมคลังข้อสอบ
พนัส จันทร์เปล่ง

ผู้ช่วยนักวิจัยทำหน้าที่บันทึกข้อมูล

ธิดาพร ไตสดี

พงษ์ลิขิต เพชรผล

อุทุมพร ชาติเผือก



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย