

การศึกษาตัวประกอบในค่าประจำข้อสอบของแบบสอบปลายภาครายวิชา
“การวัดและประเมินผลการเรียนรู้” เมื่อวิเคราะห์ข้อสอบรายข้อโดยวิธีต่างกัน 4 วิธี
Factors in the item analysis indices in the Final Test for Measurement
and Evaluation Course using 4 Difference Methods.

พิศิษฐ์ ตัณฑวณิช³

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อหาค่าประจำข้อสอบของแบบสอบปลายภาครายวิชา การวัดและประเมินผลการเรียนรู้ เมื่อวิเคราะห์โดยวิธีการต่างกัน 4 วิธี คือการหาค่าความยากและค่าอำนาจจำแนกในแนวทฤษฎีดั้งเดิม การหาค่าความยาก ค่าอำนาจจำแนกและค่าการเดาตามทฤษฎีการตอบสนองต่อข้อสอบ ค่าความคลาดเคลื่อนของข้อสอบกรณีผู้รอบเป็นผู้รอบรู้ตอบผิด (ค่า-s) ค่าความคลาดเคลื่อนของข้อสอบกรณีผู้สอบที่เป็นผู้ไม่รอบรู้ตอบถูก (ค่าการเดา-g) และค่าอำนาจจำแนก (ค่า IDI) โดยใช้ The DINA Model และ The DINO Model ตามแนวทฤษฎีการพัฒนาแบบสอบวินิจัยโดยวิธีการทางพุทธิปัญญา หาค่าความสัมพันธ์ระหว่างค่าประจำข้อสอบเมื่อวิเคราะห์โดยวิธีต่างๆ และตัวประกอบของค่าประจำข้อสอบเมื่อวิเคราะห์โดยวิธีต่างกัน

แบบสอบที่ใช้ประกอบการวิจัยเป็นแบบสอบปลายภาครายวิชา การวัดและประเมินผลการเรียนรู้ ที่คณาจารย์ในโปรแกรมทดสอบและวิจัยการศึกษา 5 ท่านร่วมกันสร้างขึ้น มีทั้งสิ้น 98 ข้อ มุ่งวัดในคุณลักษณะหลักรวม 9 ประการ กลุ่มตัวอย่างที่ใช้ในการศึกษาเป็นนักศึกษาสายครุศาสตรบัณฑิต ในภาคเรียนที่ 2 ปีการศึกษา 2560 มหาวิทยาลัยราชภัฏสุรินทร์ รวมทั้งสิ้น 968 คน ผลการวิเคราะห์โดยสรุปดังนี้

1. ข้อสอบ 98 ข้อ มีค่า p เฉลี่ยเป็น 0.45 ส่วนเบี่ยงเบนมาตรฐานเป็น 0.20 มีค่ามัธยฐานเป็น 0.44 มีค่าต่ำสุดเป็น 0.09 มีค่าสูงสุดเป็น 0.91 มีข้อสอบที่มีค่าอำนาจจำแนก 0.20 หรือสูงกว่า รวม 61 ข้อ มีความเชื่อมั่นโดยการหาค่า KR20 เป็น 0.82

2. ผลการศึกษาความสัมพันธ์สหสัมพันธ์ของค่าประจำข้อสอบเมื่อวิเคราะห์โดยวิธีการต่างกันพบว่า ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าความยาก ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าอำนาจจำแนก และค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าการเดา มีค่าสูงอย่างมีนัยสำคัญทางสถิติที่ระดับ .01

3. ผลการวิเคราะห์ตัวประกอบของค่าประจำข้อสอบเมื่อวิเคราะห์ค่าประจำข้อสอบโดยวิธีการต่างกัน ใช้วิธีการสกัดตัวประกอบแบบ Unweighted Least Square (ULS) หมุนแกนโดยวิธี Varimax ได้ตัวประกอบที่สำคัญ 2 ตัว คือ ตัวประกอบตัวแรกคือตัวประกอบ “ค่าความยากที่ฟังก์ชันการเดา” ประกอบด้วยค่าประจำข้อสอบคือค่าความยากตามแนวทฤษฎีดั้งเดิม ค่าความยากและค่าการเดาตามแนวทฤษฎีการตอบสนองต่อข้อสอบ ค่า $1 - s$ และค่า g ตามแนวทฤษฎีการพัฒนาแบบสอบวินิจัยโดยวิธีการทางพุทธิปัญญา ตัวประกอบตัวที่สองคือ ตัวประกอบด้านค่าอำนาจจำแนก ประกอบด้วยค่าอำนาจจำแนกของข้อสอบตามแนวทฤษฎีดั้งเดิม ค่าอำนาจจำแนกตามแนวทฤษฎีการตอบสนองต่อข้อสอบ และค่า IDI ตามแนวทฤษฎีการพัฒนาแบบสอบวินิจัยโดยวิธีการทางพุทธิปัญญา



Abstract

The objectives of this study were to developed and find item indices, to find correlation coefficients amongst item indices and to find factors amongst item indices of the final test for measurement and evaluation course. The item analysis techniques used consisted of the classical test theory, the item response theory, and the DINA Model and The DINO Model from the cognitive diagnostic assessment theory. The 98 items-test measured 9 attributes were developed by 5 instructors in Measurement and Evaluation Program, Faculty of Education, Surin Rajabhat University. Samples consisted of 968 students studying in the third year for Bachelor of Education Degrees, and took the course in the second semester, 2017 academic year. The finding were as the following :

1. The average of items difficulties was 0.45, which SD. "was" 0.20.

The minimum p index was 0.09 and the maximum was 0.91. The median was 0.44. The items which had their discrimination indices as 0.20 or higher were 61 items.

The reliability index computed by the KR20 Formula was 0.82

2. The correlation coefficients amongst difficulty indices, item discrimination indices and item guessing indices were statistically significant at 0.01 level

3. Two Factors were found from item indices when The Unweighted Least Square Method to extract the factor and varimax rotation method was used. The first factor was named "the item difficulty and added-guessing factor. It consisted of p index, b index, c index, 1- s, and g indices. The second factor was named "the item discrimination factor." It consisted of r, a and IDI indices.

บทนำ

ในภาวะปัจจุบัน ทฤษฎีการสร้างและพัฒนาแบบสอบมีความก้าวหน้าและเปลี่ยนแปลงไปมาก รวมทั้งโปรแกรมเพื่อการวิเคราะห์ข้อสอบในวิธีการวิเคราะห์แบบใหม่ๆ ก็ได้รับการพัฒนาขึ้นด้วยเช่นกัน ในช่วงต้นๆ ของการพัฒนา โปรแกรมการวิเคราะห์ข้อสอบที่มีใช้ คือโปรแกรมคือ การวิเคราะห์ข้อสอบในแนวทฤษฎีการวัดผลแบบดั้งเดิม (Classical Test Theory ตัวย่อ CTT) โดยมุ่งหาค่าความยาก ค่าอำนาจจำแนก ต่อมาเกิดแนวคิดวิธีการวิเคราะห์ตามแนวทฤษฎีการตอบสนองต่อข้อสอบ (Item Response Theory ตัวย่อ IRT) โดยมุ่งหาค่าประจำข้อสอบ 3 ค่าคือ ค่าอำนาจจำแนก (a) ค่าความยาก (b) และค่าการเดา (c) และต่อมามีการพัฒนาทฤษฎีการพัฒนาแบบสอบวินิจฉัยโดยวิธีการทางพุทธิปัญญา (Cognitive Diagnostic Model ตัวย่อ CDM) โดยเริ่มต้นจากงานของ Tatsuoka, K.K. (1984) และ Tatsuoka, K.K. and Tatsuoka, M.M. (1992) ในการพัฒนาแบบสอบวินิจฉัยโดยวิธีการทางพุทธิปัญญาปรากฏผลของงานวิจัยที่ดำเนินการโดย Lee, de la torre and Park (2012) โดยใช้แบบสอบ The Florida Comprehensive Assessment Test (FACT) จำนวน 50 ข้อในส่วนของวิชาคณิตศาสตร์ กำหนดคุณลักษณะที่มุ่งวัดรวม 13 คุณลักษณะ กลุ่มตัวอย่างที่ใช้เป็นนักเรียนระดับเกรด 9 ทั้งหมด 1,500 คน ผลการศึกษาพบว่า ค่าความยากในแนวทฤษฎีดั้งเดิม (p) ค่าความยากในแนวทฤษฎีการตอบสนองต่อข้อสอบ (b) และค่าโอกาส



ความน่าจะเป็นในการตอบข้อสอบถูกของผู้สอบที่มีความรอบรู้ในทุกคุณลักษณะที่ต้องใช้การตอบข้อสอบให้ถูกต้องตามแนว The DINA Model (โมเดลย่อยส่วนหนึ่งของทฤษฎี CDM) มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติ ทั้งพบว่าค่าความยาก ค่าอำนาจจำแนกของวิธีการวิเคราะห์ข้อสอบทั้ง 3 วิธี รวมทั้งค่าการเดาตอบ ถูกข้อสอบข้อสอบตามแนวทฤษฎีการตอบสนองต่อข้อสอบและในแนว The DINA Model แต่ละกลุ่มมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติ

ซึ่งจากการเกิดมีแนวทฤษฎีการวิเคราะห์ข้อสอบเกิดขึ้นหลาย ๆ แนว นั้น แม้จะมีข้อดีคือก่อให้เกิดความก้าวหน้าและความหลากหลายทางวิชาการ แต่ก็แฝงไว้ด้วยปัญหาหลัก ๆ ทางวิชาการเช่นกัน คือ มีแนวคิดการวิเคราะห์ข้อสอบที่มากมายหลากหลายจนเกินไป จนอาจนำไปสู่การยากที่จะทำความเข้าใจและเลือกใช้ งานของผู้ปฏิบัติงานขั้นพื้นฐาน ทั้งนี้หากมีการตรวจสอบว่า แนวการวิเคราะห์ข้อสอบที่ปรากฏในภาคปฏิบัติ อย่างหลากหลายนั้น โดยเบื้องต้นแล้วร่วมวัดคุณลักษณะภายในตัวมนุษย์ได้ไม่มากมายดังที่ปรากฏเป็นค่าการวิเคราะห์ข้อสอบที่ต่างนักคิดต่างเสนอออกมา

ก็จะเป็นการจัดระบบ รวมและกระชับความคิดที่แตกกระจายให้เป็นกลุ่มก้อน และเป็นการทำทางไปสู่งานการวิเคราะห์ข้อสอบที่ชัดเจน และสื่อความหมายเบื้องต้นได้ตรงกัน

นอกจากนี้ในปี 2560 จากผลการวิจัยของคณาจารย์โปรแกรมทดสอบและวิจัยการศึกษา คณะครุศาสตร์ มหาวิทยาลัยราชภัฏสุรินทร์ ประกอบด้วย เกื้อ กระแสโสม พิเศษฐ ตันทวนิช พนา จินดาศรี ตรีคม พรหมมาบุญ และ บุญทวี อิมบุญตา (2560) เมื่อศึกษากับแบบสอบปลายภาครายวิชาการวิจัยเพื่อพัฒนาการเรียนรู้อย่างเป็นวิชาชีพของนักศึกษาสายครุศาสตร์บัณฑิต มหาวิทยาลัยราชภัฏสุรินทร์ ปีการศึกษา 2561 จำนวน 1070 คน เมื่อนำข้อสอบดังกล่าวจำนวน 100 ข้อมาวิเคราะห์ในลักษณะเดียวกันกับงานของ Lee, de la torre and Park ปรากฏผลในเบื้องต้นว่าเป็นไปในทิศทางเดียวกันกับงานวิจัยที่ได้กล่าวถึง

อย่างไรก็ดี การเกิดมีแนวทฤษฎีการวิเคราะห์ข้อสอบเกิดขึ้นหลาย ๆ แนว นั้น แม้จะมีข้อดีคือก่อให้เกิดความก้าวหน้าและความหลากหลายทางวิชาการ แต่ก็แฝงไว้ด้วยปัญหาหลัก ๆ ทางวิชาการเช่นกัน กล่าวคือการมีแนวคิดการวิเคราะห์ข้อสอบที่มากมายหลากหลายจนเกินไป จนอาจนำไปสู่การยากที่จะทำความเข้าใจและเลือกใช้ งานของผู้ปฏิบัติงานขั้นพื้นฐาน รวมทั้งหากมีการตรวจสอบว่า แนวการวิเคราะห์ข้อสอบที่ปรากฏในภาคปฏิบัติอย่างหลากหลายนั้น โดยเบื้องต้นแล้วร่วมวัดคุณลักษณะภายในตัวมนุษย์ได้ไม่มากมายดังที่ปรากฏเป็นค่าการวิเคราะห์ข้อสอบที่ต่างนักคิดต่างเสนอออกมา ก็จะนำไปสู่แนวคิดใหม่ของการวิจัยในด้านการจัดระบบ รวมและกระชับความคิดที่แตกกระจายมากมายให้เป็นกลุ่มก้อน และเป็นการทำทางไปสู่งานการวิเคราะห์ข้อสอบที่ชัดเจน ให้ความหมายและสื่อความหมายเบื้องต้นได้ตรงกัน

วัตถุประสงค์ของการวิจัย

ในการวิจัยครั้งนี้มีวัตถุประสงค์ดังนี้

1. เพื่อหาค่าประจำข้อสอบของแบบสอบปลายภาคในรายวิชาการวัดและประเมินผลการเรียนรู้เมื่อวิเคราะห์โดยวิธีการต่างกัน 4 วิธีดังนี้

- วิเคราะห์หาค่าความยาก (p) และค่าอำนาจจำแนกของข้อสอบ (r) ตามแนวคิดทฤษฎีดั้งเดิม
- วิเคราะห์หาค่าอำนาจจำแนก (a) ค่าความยาก (b) และค่าการเดา (c) เมื่อวิเคราะห์ข้อสอบ

โดยทฤษฎีการตอบสนองต่อข้อสอบ



- วิเคราะห์หาค่าโอกาสความน่าจะเป็นในส่วนของความคลาดเคลื่อนของผู้รอบรู้ในทุกคุณลักษณะที่ต้องใช้ในการตอบข้อสอบข้อหนึ่งตอบผิดในข้อสอบข้อนั้นๆ (ค่า s) และเพื่อหาค่าโอกาสความน่าจะเป็นในส่วนของความคลาดเคลื่อนที่ผู้ไม่รอบรู้ในทุกคุณลักษณะที่ต้องใช้ในการตอบข้อสอบข้อหนึ่งๆ สามารถตอบข้อสอบข้อนั้นได้ถูกต้อง (ค่าการเดา g) และหาค่าอำนาจจำแนก (IDI) ของข้อสอบรายข้อเมื่อวิเคราะห์ข้อสอบโดยวิธีการสร้างแบบสอบวินิจฉัยโดยวิธีการทางพุทธิปัญญา โดยใช้ The DINA Model และ The DINO Model

2. เพื่อค่าความสัมพันธ์ระหว่างค่าประจำข้อสอบในแบบสอบเมื่อวิเคราะห์โดยวิธีการต่างกันทั้ง 4 วิธีในวัตถุประสงค์ข้อ 1

3. เพื่อหาตัวประกอบของค่าประจำข้อสอบจากผลการวิเคราะห์ข้อสอบโดยวิธีการวิเคราะห์ข้อสอบต่างกัน 4 วิธีตามวัตถุประสงค์ข้อ 1

ขอบเขตของการวิจัย

การวิจัยครั้งนี้มีขอบเขตดังนี้

1. ประชากร ในการวิจัยครั้งนี้หมายถึงนักศึกษาสายครุศาสตรบัณฑิตในระดับชั้นปีที่ 3 ซึ่งต้องเรียนวิชา “การวัดและประเมินผลการเรียนรู้” เป็นรายวิชาซีพีครุบังคับ กลุ่มตัวอย่างหมายถึงนักศึกษาสายครุศาสตร์ ระดับชั้นปีที่ 3 สังกัดมหาวิทยาลัยราชภัฏสุรินทร์ที่เรียนวิชา “การวัดและประเมินผลการเรียนรู้” ในภาคเรียนที่ 2 ปีการศึกษา 2560 รวม 968 คน

2. การหาค่าอำนาจจำแนกของข้อสอบในแบบสอบรายวิชาการวิจัยเพื่อพัฒนาการเรียนรู้เกี่ยวกับการวิจัย ใช้วิธีการหาค่า point biserial correlation ของข้อสอบข้อนั้น กับค่าคะแนนรวมของแบบสอบทั้งฉบับ

3. การหาค่าประจำข้อสอบตามแนวทฤษฎีการตอบสนองต่อข้อสอบใช้โปรแกรมวิเคราะห์ข้อมูลที่มุ่งหาค่าประจำข้อสอบทั้ง 3 พารามิเตอร์ คือ อำนาจจำแนก ค่าความยาก และค่าการเดา

4. การหาค่าประจำข้อสอบตามแนวการสร้างแบบสอบวินิจฉัยโดยวิธีการทางพุทธิปัญญา มุ่งศึกษาเฉพาะการหาค่าประจำข้อสอบในแนว The DINA Model และ The DINO Model เท่านั้น

สมมุติฐานของการวิจัย

ในการวิจัยครั้งนี้มีสมมุติฐานสำคัญดังนี้

1. ในการศึกษาความสัมพันธ์ระหว่างค่าประจำข้อสอบข้อสอบเมื่อวิเคราะห์โดยวิธีต่างกัน 4 วิธีจะ
ได้ผลดังนี้

1.1 ค่าความยากของข้อสอบเมื่อวิเคราะห์โดยวิธีการต่างกัน 4 วิธี มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติ

1.2 ค่าอำนาจจำแนกของข้อสอบเมื่อวิเคราะห์โดยวิธีการต่างกัน 4 วิธี มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติ

1.3 ค่าการเดาของข้อสอบเมื่อวิเคราะห์ตามแนวทฤษฎีการสร้างแบบสอบวินิจฉัยโดยวิธีการทางพุทธิปัญญา (g) และค่าการเดา (ค่า c) ของข้อสอบตามแนวทฤษฎีการตอบสนองต่อข้อสอบมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติ



2. เมื่อนำค่าเมตริกสหสัมพันธ์ระหว่างค่าประจำข้อสอบมาวิเคราะห์โดยวิธีการวิเคราะห์ตัวประกอบ จะได้ตัวประกอบ 3 ตัวคือ

2.1 ตัวประกอบของค่าความยาก จะสอดคล้องกับสมมุติฐานข้อ 1.1 คือเป็นตัวประกอบด้านค่าความยาก

2.2 ตัวประกอบของค่าอำนาจจำแนก จะสอดคล้องกับสมมุติฐานข้อ 1.2 คือเป็นตัวประกอบด้านค่าอำนาจจำแนกของข้อสอบ

2.3 ตัวประกอบของค่าการเดา จะสอดคล้องกับสมมุติฐานข้อ 1.3 คือเป็นตัวประกอบด้านค่าการเดา

ทฤษฎีที่เกี่ยวข้อง

เนื่องจากในงานวิจัยครั้งนี้มีการนำเอาแนวคิดทฤษฎีการพัฒนาข้อสอบแนวใหม่ที่เรียกว่า การพัฒนาแบบสอบวินิจฉัยโดยวิธีการทางพุทธิปัญญา (Cognitive Diagnostic Model) มาใช้ จึงขอนำเสนอแนวคิดของทฤษฎีดังกล่าวบางส่วน กล่าวคือทฤษฎีนี้นำเสนอว่าข้อสอบที่จะใช้ได้ดีในการที่จะระบุถึงจุดบกพร่องในการเรียนรู้ของผู้เรียนนั้น ผู้สร้างข้อสอบจะต้องระบุให้ได้ว่า ข้อสอบข้อหนึ่งๆ นั้นมุ่งวัดคุณลักษณะ (Attribute) ไตบ้าง โดยคำว่าคุณลักษณะในที่นี้หมายถึงสมรรถภาพทางสมองของผู้เรียนที่จำเป็นต้องใช้เพื่อการหาคำตอบที่ถูกต้องให้กับข้อสอบแต่ละข้อในแบบสอบฉบับหนึ่ง ๆ ทฤษฎีดังกล่าวนี้ได้รับการพัฒนาขึ้นอย่างมากในช่วงยี่สิบปีที่ผ่านมา จนเกิดเป็นโมเดลการสร้างและวิเคราะห์ข้อสอบขึ้นหลายโมเดลในที่นี้ขอกล่าวถึงโมเดลส่วนที่นำมาใช้ในงานวิจัยครั้งนี้คือ

-The DIN A Model แนวคิดของโมเดลนี้อธิบายว่า ในการตอบข้อสอบข้อหนึ่ง ๆ ให้ถูกต้องนั้น ผู้สอบจะต้องใช้ความสามารถย่อยๆ ทุกๆ คุณลักษณะ (Attribute) จำนวนหนึ่งมาช่วยกัน เพื่อให้ทำข้อสอบข้อดังกล่าวนั้น ๆ ได้ถูกต้อง แนวคิดนี้มุ่งหาค่าประจำข้อสอบ 2 ค่าหลัก คือ

ค่า s หมายถึงค่าโอกาสความน่าจะเป็นของความคลาดเคลื่อนที่จะเกิดขึ้น เมื่อผู้สอบมีความรู้ในทุกคุณลักษณะย่อยที่ต้องใช้ในการตอบข้อสอบข้อดังกล่าวให้ถูกต้อง แต่ผู้สอบคนดังกล่าวกลับตอบผิด

ค่า $1 - s$ หมายถึงค่าความยากของข้อสอบเมื่อวิเคราะห์ข้อสอบตามแนวทฤษฎีการสร้างแบบสอบวินิจฉัยโดยวิธีการทางพุทธิปัญญา

ค่า g หมายถึงค่าการเดา หรือคือค่าของโอกาสความน่าจะเป็นของความคลาด-เคลื่อนที่จะเกิดขึ้น เมื่อผู้สอบมีความรู้ไม่ครบทุกคุณลักษณะย่อยที่ต้องใช้ในการตอบข้อสอบข้อดังกล่าวให้ถูกต้อง แต่ผู้สอบคนดังกล่าวกลับตอบถูก

ค่า IDI หมายถึงค่าอำนาจจำแนกของข้อสอบแต่ละข้อตามการสร้างแบบสอบวินิจฉัยโดยวิธีการทางพุทธิปัญญา โดยค่าดังกล่าวคำนวณได้จากสมการคือ $IDI = 1 - s - g$

-The DINO Model แนวคิดของโมเดลนี้อธิบายว่า ในการตอบข้อสอบข้อหนึ่ง ๆ ให้ถูกต้องนั้น ผู้สอบจะต้องใช้ความสามารถย่อย ๆ ที่เรียก “คุณลักษณะ” (Attribute) จำนวนหนึ่งมาช่วยกัน เพื่อให้ทำข้อสอบข้อดังกล่าวนั้น ๆ ได้ถูกต้อง โดยมุ่งหาค่าประจำข้อสอบ 2 ค่าหลัก เช่นเดียวกับ The DIN A Model แต่มีความแตกต่างตรงที่ว่า The DINO Model นั้น ถือว่า การที่ผู้สอบจะทำข้อสอบข้อนั้น ๆ ได้ถูกต้อง เขาอาจจะมีความรู้เพียงบางคุณลักษณะ จากคุณลักษณะทั้งหมดที่กำหนด หรือรู้เพียงคุณลักษณะเดียวก็สามารถทำให้ผู้สอบตอบข้อสอบข้อดังกล่าวได้ถูกต้อง



วิธีดำเนินการวิจัย

การพัฒนาแบบสอบ

ขั้นตอนที่สำคัญที่สุดของการวิจัยด้านการพัฒนาแบบสอบในแนวการพัฒนาแบบทดสอบวินิจฉัยโดยวิธีการทางพุทธิปัญญา คือการกำหนดคุณลักษณะที่มุ่งวัดในแบบสอบ โดยใน”แบบสอบแนวคิดพื้นฐานการวิเคราะห์ข้อสอบแบบ” ผู้วิจัยได้ดำเนินการสร้างแบบสอบตามหลักการของการสร้างแบบสอบวินิจฉัยโดยวิธีการทางพุทธิปัญญา เริ่มต้นจากกำหนดคุณลักษณะมุ่งศึกษาในแบบสอบรวม 9 ข้อดังนี้

1. ความรู้เบื้องต้นเกี่ยวกับการวัดและประเมินผล ใช้ข้อสอบจำนวน 11 ข้อ
2. หลักการ แนวคิดหรือประเภทของการวัดและประเมิน ใช้ข้อสอบจำนวน 12 ข้อ
3. ประเภทของพฤติกรรมทางการศึกษา ใช้ข้อสอบจำนวน 14 ข้อ
4. แบบสอบวัดผลสัมฤทธิ์ทางการเรียน ใช้ข้อสอบจำนวน 11 ข้อ
5. การประเมินผลภาคปฏิบัติ (Performance assessment) ใช้แฟ้มสะสมงาน ใช้ข้อสอบจำนวน 17 ข้อ
6. การวิเคราะห์ข้อสอบรายข้อ ใช้ข้อสอบจำนวน 13 ข้อ
7. ความตรงและความเที่ยงของแบบสอบ ใช้ข้อสอบจำนวน 18 ข้อ
8. สถิติที่จำเป็นต้องนำมาใช้ในการวัดผลประเมินผล ใช้ข้อสอบจำนวน 12 ข้อ
9. การตัดสินผลการเรียนรู้ ใช้ข้อสอบจำนวน 8 ข้อ
(มีข้อสอบบางข้อที่วัดคุณลักษณะมากกว่า 1 คุณลักษณะ)

การตรวจสอบความสามารถของข้อสอบที่วัดได้ตรงตามคุณลักษณะทั้ง 9 ข้อที่กำหนด

การตรวจสอบว่าข้อสอบแต่ละข้อที่สร้างขึ้นมานั้นวัดคุณลักษณะที่มุ่งศึกษาในข้อใดบ้าง กระทำโดยการพิจารณาจากผู้วิจัยร่วมกันทั้ง 5 ท่าน ทั้งนี้การพิจารณากระทำโดยการลงความเห็นที่เป็นอิสระของแต่ละคน และการตัดสินผลว่าข้อสอบสอบแต่ละข้อวัดในคุณลักษณะใดบ้าง จะต้องผ่านความเห็นร่วมกัน 3 ใน 5 ท่านขึ้นไป และนำเอาผลดังกล่าวมากำหนดเป็นค่า Q-Matrix

เพื่อบรรยายความสัมพันธ์ระหว่างข้อสอบแต่ละข้อกับคุณลักษณะทั้ง 9 ข้อที่มุ่งวัด

การเก็บข้อมูล

การเก็บรวบรวมข้อมูล ดำเนินการตามกำหนดการสอบปลายภาคเรียนในภาคเรียนที่ 2 ปีการศึกษา 2560 โดยตกในปลายเดือนกุมภาพันธ์ 2561 ภาพรวมคะแนนจากข้อสอบ 98 ข้อมีค่าเฉลี่ยเป็น 44.07 คิดเป็นร้อยละ 44.97 ส่วนเบี่ยงเบนมาตรฐานเป็น 10.49 คะแนนมัธยฐานเป็น 43 คะแนนต่ำสุด 15 คิดเป็นร้อยละ 15.31 คะแนนสูงสุด 80 คะแนน คิดเป็นร้อยละ 81.63 มีความเชื่อมั่นโดยการหาค่า KR20 เป็น 0.82 ค่าอำนาจจำแนกในรูปของ Point Biserial Correlation โดยเฉลี่ยมีค่า 0.26

การวิเคราะห์ข้อมูล

การวิเคราะห์ข้อมูลมีขั้นตอนที่ดำเนินการดังนี้

1. การหาค่าความยาก ค่าอำนาจจำแนกของข้อสอบ และค่าความเชื่อมั่นของแบบสอบโดยการหาค่า KR20 ใช้วิธีการวิเคราะห์โดยใช้โปรแกรมสำเร็จรูปคือโปรแกรม TAP (Test Analysis Program) ซึ่งพัฒนาโดย Gordon P. Brooks (2002)
2. การหาค่าพารามิเตอร์ทั้งสามค่าตามแนวทฤษฎีการตอบสนองต่อข้อสอบ ใช้โปรแกรมสำเร็จรูป EXCALIBRE (Assessment Systems Corporation, 1967) ในการวิเคราะห์



3. การหาค่า s และค่า g เมื่อวิเคราะห์ข้อสอบโดยวิธีการสร้างแบบสอบวินิจฉัยโดยวิธีการทางพุทธิปัญญา โดยใช้ The DINA Model และ The DINO Model ใช้โปรแกรมที่มีชื่อว่า The CDM Package ซึ่งทำงานบนโปรแกรม Program R พัฒนาโดย Robitzsch, A and others (2015) เป็นโปรแกรมวิเคราะห์ข้อมูล

4. การหาค่าความสัมพันธ์ของค่าประจำข้อสอบที่วิเคราะห์โดยวิธีการต่างกัน ใช้วิธีการหาค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน

5. การวิเคราะห์ตัวประกอบ ใช้วิธีการสกัดน้ำหนักตัวประกอบโดยวิธี Unweighted Least Square (ULS) หมุนแกนโดยวิธี Varimax

ผลการวิเคราะห์ข้อมูล

ผลการวิเคราะห์ข้อมูล สามารถเสนอได้เป็นตารางหลัก 3 ตารางดังต่อไปนี้
ตารางที่ 1 แสดงค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ค่ามัธยฐานค่าสูงสุด ค่าต่ำสุด ของค่าประจำข้อสอบของข้อสอบแต่ละข้อเมื่อวิเคราะห์โดยวิธีการต่างกัน 4 วิธี

แนวการวิเคราะห์	ค่าประจำข้อสอบ	ค่าเฉลี่ย	ส่วนเบี่ยงเบนมาตรฐาน	ค่ามัธยฐาน	ค่าต่ำสุด	ค่าสูงสุด
CTT Model	"p"	0.45	0.20	0.44	0.09	0.91
	"r"	0.23	0.10	0.23	0.02	0.45
	"a"	0.48	0.13	0.45	0.26	0.90
IRT Model	"b"	1.03	1.49	1.06	-2.80	3.00
	"c"	0.17	0.02	0.17	0.10	0.20
	"s"	0.57	0.22	0.58	0.11	0.97
The DINO Model	"g"	0.36	0.19	0.34	0.05	0.85
	"IDI"	0.21	0.13	0.20	-0.02	0.64
	"s"	0.58	0.22	0.58	0.11	0.97
The DINA Model	"g"	0.36	0.19	0.34	0.0528	0.85
	"IDI"	0.21	0.13	0.20	-0.02	0.67

จากตารางที่ 1 โดยภาพรวม แสดงว่าข้อสอบที่พัฒนาขึ้น ค่อนข้างยากและอำนาจจำแนกไม่สูงนัก เมื่อวิเคราะห์ตามแนวคิดทฤษฎีดั้งเดิม แนวการพัฒนาแบบสอบวินิจฉัยโดยวิธีการทางพุทธิปัญญา และแนวทฤษฎีการตอบสนองต่อข้อสอบ



ตารางที่ 2 แสดงค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่า p r ค่า a b และ c ค่า 1- s g และ IDI ของข้อสอบ 98 ข้อที่ได้พัฒนาขึ้น

ค่า สถิติ	CTT Model		IRT Model			The DINA Model			The DINO Model		
	"p"	"r"	"a"	b	c	1 - s	"g"	IDI	1- s	"g"	IDI
"p"											
"r"	0.24*										
A	-0.53**	0.47**									
B	-0.97**	-0.35**	0.36**								
C	0.57	-0.24*	-0.83**	0.38**							
1- s(DINA)	0.93**	0.55**	-0.28**	-0.94**	0.42**						
"g"(DINA)	0.96**	0.01	-0.64**	-0.92**	0.62**	0.82**					
"IDI(DINA)	0.15	0.94**	0.48**	-0.25*	-0.21*	0.50**	-0.10				
1- s(DINO)	0.92**	0.55**	-0.28**	-0.94*	0.42*	0.99**	0.80**	0.51**			
"g"(DINO)	0.97**	0.01	-0.64	-0.92**	0.62**	0.82**	0.99**	-0.09	0.81**		
"IDI(DINO)	0.15	0.93**	0.47**	-0.25*	-0.21*	0.49**	-0.10	0.99**	0.51**	-0.09	

เมื่อพิจารณาข้อมูลในตารางที่ 2 ในเบื้องต้น จะพบว่า ค่าความยากในแนวทฤษฎีดั้งเดิม ค่าความยากในแนวทฤษฎีการตอบสนองต่อข้อสอบ ค่าความยากตามแนวการพัฒนาแบบสอบวินิจัยโดยวิธีการทางพุทธิปัญญา มีค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างกันสูงอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01

ค่าอำนาจจำแนกในแนวทฤษฎีดั้งเดิม ค่าอำนาจจำแนกในแนวทฤษฎีการตอบสนองต่อข้อสอบและค่าอำนาจจำแนกตามแนวการพัฒนาแบบสอบวินิจัยโดยวิธีการทางพุทธิปัญญา มีค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างกันสูงอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01

ค่าการเดาในแนวทฤษฎีการตอบสนองต่อข้อสอบกับค่าการเดาของข้อสอบตามแนวการพัฒนาแบบสอบวินิจัยโดยวิธีการทางพุทธิปัญญามีสัมประสิทธิ์สหสัมพันธ์ระหว่างกันสูงอย่างมีนัยสำคัญ ทางสถิติที่ระดับ 0.01

ตารางที่ 3 แสดงค่า Factor Loading ใน Factor ที่สกัดได้จากค่าสัมประสิทธิ์สหสัมพันธ์ ระหว่างค่าประจำข้อสอบที่วิเคราะห์ได้โดยวิธีต่างกัน 4 วิธี ของข้อสอบ 98 ข้อ

ค่าประจำข้อสอบ	ตัวประกอบที่ 1	ตัวประกอบที่ 2
"p"	0.99	0.12
"r"	0.12	0.96
"a"	-0.62	0.58
"b"	-0.92	-0.25
"c"	0.63	-0.33
1- s (DINA)	0.89	0.47



“g” (DINA)	0.98	-0.12
IDI (DINA)	0.05	0.97
1- s (DINO)	0.88	0.47
“g” (DINO)	0.99	-0.12
IDI (DINO)	0.05	0.96
Eigen Value	6.27	3.63
SS Loadings	55.74	34.23
Cu SS Loadings	55.74	89.97

จากตารางที่ 3 ผลการวิเคราะห์ตัวประกอบพบว่า ค่า Factor Loading ของค่าความยากในแนวทฤษฎีดั้งเดิม ค่าความยากในแนวทฤษฎีการตอบสนองต่อข้อสอบ ค่าความยากตามแนวการสร้างแบบสอบวินิจฉัยโดยวิธีการทางพุทธิปัญญา ค่าการเดาตามแนวทฤษฎีการตอบสนองต่อข้อสอบ และค่าการเดาของข้อสอบตามแนวการสร้างแบบสอบวินิจฉัยโดยวิธีการทางพุทธิปัญญา รวมตัวกันเป็นตัวประกอบตัวที่ 1 โดยมีค่าไอเก้นที่ระดับ 6.27 ให้ชื่อว่า “ตัวประกอบค่าความยากที่ฟุ้งค่าการเดา” และพบว่าค่า Factor Loading ของค่าความอำนาจจำแนกในแนวทฤษฎีดั้งเดิม ค่าอำนาจจำแนกในแนวทฤษฎีการตอบสนองต่อข้อสอบ ค่าอำนาจจำแนกตามแนวการสร้างแบบสอบวินิจฉัยโดยวิธีการทางพุทธิปัญญา รวมตัวกันเป็นตัวประกอบตัวที่ 2 โดยมีค่าไอเก้นที่ระดับ 3.63 ให้ชื่อว่า “ตัวประกอบค่าอำนาจจำแนก” (โดยตัวประกอบจากตัวที่ 3 เป็นต้นไปมีค่าไอเก้นต่ำกว่า 1) โดยตัวประกอบ 2 ตัวร่วมกันอธิบายความแปรปรวนได้ร้อยละ 89.97 จากความแปรปรวนทั้งหมด

สรุปผลการวิจัย

ผลสำคัญที่ได้จากวิจัยครั้งนี้มีดังนี้

1. ค่าประจำข้อสอบที่วิเคราะห์ได้ทั้งเมื่อใช้แนวคิดทฤษฎีดั้งเดิม ทฤษฎีการตอบสนองต่อข้อสอบ และการวิเคราะห์ตามวิธีการสร้างแบบสอบวินิจฉัยโดยวิธีการทางพุทธิปัญญา แสดงว่าโดยรวมข้อสอบที่พัฒนาขึ้นค่อนข้างยากและอำนาจจำแนกไม่สูงนัก
2. ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าความยาก ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าอำนาจจำแนก และค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าการเดา ที่วิเคราะห์โดยวิธีการต่างกันมีค่าสูงอย่างมีนัยสำคัญทางสถิติที่ระดับ .01
3. ตัวประกอบที่วิเคราะห์มี 2 ตัวประกอบหลัก คือตัวประกอบด้านค่าความยากที่ฟุ้งค่าการเดา ส่วนนี้ไม่เป็นไปตามสมมุติฐานที่กำหนด และตัวประกอบด้านค่าอำนาจจำแนก ซึ่งเป็นส่วนที่เป็นไปตามสมมุติฐาน



อภิปรายผลการวิจัย

ในด้านการอภิปรายผล มีสาระที่ควรอภิปรายรวม 3 ประเด็นดังนี้

1. ผลจากการวิเคราะห์ค่าประจำข้อสอบโดยวิธีการต่างกัน แต่ให้ผลการแปลความหมายโดยรวมร่วมกันว่า ข้อสอบในแบบสอบที่พัฒนาขึ้นมีคุณภาพอยู่ในเกณฑ์ที่ใช้ได้ในเบื้องต้นแต่ค่อนข้างยาก และมีอำนาจจำแนกไม่สูงมากนัก ส่วนนี้ถือได้ว่าเป็นข้อค้นพบที่ดีที่จะช่วยชี้แนวทางในการเขียนข้อคำถามระยะต่อไปว่า ข้อสอบที่ควรพัฒนานั้นไม่ควรที่จะเน้นตรงความยาก แต่ควรหันมาเน้นการทำให้ข้อสอบมีอำนาจจำแนกสูงเป็นส่วนสำคัญ

2. ผลการศึกษาความสัมพันธ์สหสัมพันธ์ของค่าประจำข้อสอบเมื่อวิเคราะห์โดยวิธีการต่างกัน พบว่า กลุ่มค่าความยาก ค่าอำนาจจำแนก และค่าการเดาที่วิเคราะห์โดยวิธีการต่างกัน พบว่ามีความสัมพันธ์กันสูงอย่างมีนัยสำคัญทางสถิติ นั่น เป็นข้อมูลเบื้องต้นที่น่าจะสะท้อนว่า แนวคิดวิธีการวิเคราะห์ข้อสอบโดยวิธีการต่างกันนั้น ในเบื้องต้นแล้ว สิ่งมุ่งศึกษาหรือวัดนั้นน่าจะเป็นสิ่งเดียวกันหรือสิ่งที่มีความร่วมกันอยู่ เป็นข้อคิดว่าการพัฒนาแบบสอบและการวิเคราะห์ข้อสอบในระยะต่อไป นักวัดผลประเมินผลอาจจะต้องพึงเล็งไปถึงสิ่งที่อยู่ในเบื้องต้นเหล่านี้ มากกว่าจะติดอยู่กับค่าประจำข้อสอบที่วิเคราะห์ได้ในแต่ละวิธีหรือแต่ละแนวคิดทฤษฎี

3. ผลการศึกษาด้านตัวประกอบ ส่วนนี้มีประเด็นย่อยที่ควรอภิปรายแยกได้เป็น 3 ประเด็นย่อยดังนี้

3.1 จากการศึกษาด้านตัวประกอบของค่าประจำข้อสอบเมื่อวิเคราะห์โดยวิธีการต่างกัน และพบว่า ตัวประกอบด้านค่าอำนาจจำแนกเป็นไปตามสมมติฐานที่กำหนด แต่ตัวประกอบด้านค่าความยากที่พ่วงค่าการเดา แทนที่จะแยกกันเป็นคนละตัวประกอบ ตามสมมติฐานที่กำหนด แต่ค่าประจำข้อสอบในส่วนนี้กลับรวมเป็นตัวประกอบเดียวกัน รวมทั้งมีค่าสัมประสิทธิ์สหสัมพันธ์บางส่วนที่สูงมาก (ค่า 1 - s ของ The DINA Model และ The DINO Model มีค่าสัมประสิทธิ์สหสัมพันธ์ที่ระดับ 0.99 เป็นต้น) มีข้ออภิปรายดังนี้

3.1.1 ลักษณะของเมตริกของค่าสัมประสิทธิ์สหสัมพันธ์ของข้อสอบเมื่อวิเคราะห์โดยวิธีการต่างกัน พบว่าเมตริกที่ได้มีคุณสมบัติเฉพาะที่เรียกว่า Nonpositive definite (NPD) ปรากฏการณ์ดังกล่าวในที่นี้ น่าจะเกิดจาก มีค่าประจำข้อสอบบางส่วนที่มีภาวะที่เป็นตัวแปรสัมพันธ์กันแบบซ้ำซ้อนในตัวมันเอง (Linear dependencies among the variables) ทำให้การสกัดตัวประกอบจะสามารถกระทำได้โดยเทคนิค Principal Component (PC) หรือ Unweighted Least Squares (ULS) เท่านั้น ในส่วนงานวิจัยครั้งนี้ น่าจะเกิดจากค่าความยากของข้อสอบโดยวิธีการวิเคราะห์ต่างกันมีค่าสัมประสิทธิ์สหสัมพันธ์บางส่วนที่สูงมากจนเกินไป จนเกิดภาวะเป็นตัวแปรเดียวกัน (รวมทั้งข้อค้นพบในส่วนนี้ไม่เคยปรากฏในรายงานการวิจัยที่เกี่ยวข้องกับการพัฒนาแบบสอบวินิจฉัยโดยวิธีการทางพุทธิปัญญามาก่อน)

3.1.2 ผลการวิจัยส่วนที่ไม่เป็นไปตามสมมติฐานคือในส่วนของสมมติฐานที่ว่า ตัวประกอบค่าความยาก และค่าการเดา เป็นค่าประจำข้อสอบที่อยู่กันคนละตัวประกอบ แต่ผลการวิจัยพบว่า ค่าประจำข้อสอบทั้งสองส่วนนี้ รวมตัวกันเป็นตัวประกอบตัวเดียวกัน ทำไมจึงเป็นเช่นนั้น คำอธิบายในเบื้องต้นนี้มีอยู่ 2 ประการคือ

ก) ความหมายที่แท้จริงของค่าการเดา ในแนวคิดการพัฒนาแบบสอบวินิจฉัยโดยวิธีการทางพุทธิปัญญานั้น Lee, de la Torre and Park (2012 : 335) ได้กล่าวอ้างถึงข้อเขียนของ de la Torre และ Douglas ในปี 2008 และงานของ Maris ในปี 1999 ไว้ว่า ค่าการเดา (g) ที่วิเคราะห์ได้ในแนว



การพัฒนาแบบสอบวินิจฉัยโดยวิธีการทางพุทธิปัญญาว่า ค่าที่ได้ดังกล่าว มิได้หมายความว่า ค่าโอกาสความน่าจะเป็นในการเดาตอบถูกเท่านั้น แต่หมายความรวมไปถึงการตอบถูกโดยที่ผู้สอบใช้คุณลักษณะอื่นๆ ที่มีความเป็นไปได้และไม่ปรากฏใน Q-matrix ที่กำหนดขึ้นไว้ในการวิเคราะห์ข้อสอบครั้งนั้น น่าจะเป็นเหตุผลที่ทำให้ค่าการเดาที่คำนวณได้จะมีค่าค่อนข้างสูง และไปมีความสัมพันธ์กับค่าความยากของข้อสอบในระดับที่สูงตามไปด้วย

ข) การเดากับค่าความยากผันแปรไปในทิศทางเดียวกันไม่ว่าโดยแนวทฤษฎีใดๆ ข้อสอบที่มีลักษณะเป็นข้อสอบแบบเลือกตอบ โอกาสการเดาตอบถูกจะเป็นสิ่งที่แฝง ร่วมผันแปรไปตามค่าความยากเสมออย่างเลี่ยงไม่ได้ ซึ่งหากเป็นดังนี้จริง ก็ควรจะรอดูจากผลการวิจัยที่ศึกษาในทำนองเดียวกันกับงานวิจัยทำนองนี้แต่ทำการศึกษากับอีกวิชาเรียนหลายๆ วิชา และในระดับการศึกษาที่ต่างกัน

3.2 ตัวประกอบด้านค่าอำนาจจำแนก

ผลการวิจัยในการศึกษาตัวประกอบด้านค่าอำนาจจำแนกนั้น สามารถมั่นใจได้ว่า แบบสอบฉบับหนึ่งๆ เมื่อวิเคราะห์โดยวิธีการต่างกัน อำนาจจำแนกที่ได้แทบจะเป็นสิ่งเดียวกัน และมั่นใจได้ว่า ค่าอำนาจจำแนกนั้นเป็นดัชนีที่สำคัญที่สามารถชี้ระบุงถึงประสิทธิภาพที่ดีของข้อคำถามในแบบสอบ เป็นดัชนีที่จำเป็นที่จะต้องทำให้ได้ ทั้งนี้จากข้อมูลในตารางที่ 3 ยังคงพบด้วยว่า ค่า a นั้นมีค่า Factor Loading มีค่าสูงทั้งในตัวประกอบที่ 1 และตัวประกอบที่ 2 โดยในตัวประกอบที่ 1 ให้ค่าที่ติดลบ ในที่นี้ผู้วิจัยได้คัดให้ไปอยู่ในตัวประกอบที่ 2 เพราะสอดคล้องรับกับหลักทฤษฎีมากกว่าการไปอยู่ในตัวประกอบที่ 1

3.3 ลักษณะเฉพาะของ Q-Matrix ที่ใช้ในงานวิจัยครั้งนี้

ปัญหาประการหนึ่งที่จะต้องมีการสืบค้นต่อก็คือปัญหาของลักษณะและวิธีการกำหนด Q-Matrix เพื่อใช้เป็นตัวกำกับลักษณะเฉพาะของแบบสอบแต่ละชุด งานวิจัยดังกล่าวนี้ลักษณะของ Q-Matrix ที่ใช้มีลักษณะเฉพาะมากเกินไปคือ ข้อสอบแต่ละข้อมักจะวัดคุณลักษณะใดคุณลักษณะหนึ่งเพียงประการเดียว ข้อคำถามไม่ได้รับการออกแบบให้เด่นชัด โดยมีกำหนดล่วงหน้าว่าข้อสอบแต่ละข้อ ควรวัดคุณลักษณะใดบ้าง โดยเฉพาะการออกข้อคำถามข้อหนึ่งๆ ให้มีการวัดคุณลักษณะหลายๆ ข้อพร้อมๆ กันไป และสลับคุณลักษณะที่มุ่งวัดไปมาอย่างตั้งใจ ข้อจำกัดของ Q-Matrix ดังกล่าวอาจจะเป็นสาเหตุหลักที่ทำให้การวิจัยในครั้งนี้ได้ผลดังข้อสรุปของงานวิจัย หากมีการปรับเปลี่ยนลักษณะของ Q-Matrix ให้ต่างจากงานวิจัยนี้มาหลายๆ และสร้างข้อคำถามให้ได้ตามลักษณะของ Q-Matrix ที่กำหนดขึ้นใหม่ ผลการวิจัยก็อาจจะได้ทิศทางที่ต่างออกไปได้

ข้อเสนอแนะจากการวิจัย

1. การนำผลการวิจัยไปใช้

1.1 แบบสอบที่พัฒนาขึ้นในเบื้องต้นมีคุณภาพดี สามารถที่จะนำไปใช้ในการทดสอบเพื่อจัดการเรียนการสอนในลักษณะของการสอบปลายภาคได้ ทั้งนี้ การพัฒนาแบบสอบให้มีคุณภาพดียิ่งขึ้นไปอีก ควรเน้นตรงการพัฒนาแบบสอบให้มีอำนาจจำแนกสูงขึ้นไปอีก โดยวิธีการที่จะทำได้คือ พยายามรักษาระดับความยากไว้ อย่าให้ข้อสอบยากขึ้นไปกว่านี้ โดยพยายามทำให้ตัวลวงมีความชัดเจนว่าเป็นตัวผิดทันทีสำหรับผู้สอบที่มีความรู้ในสาระที่ถามอย่างดี แต่ผู้สอบที่ไม่รู้จริง ไม่สามารถแยกแยะได้ว่า ตัวเลือกตัวใดเป็นตัวถูกและตัวเลือกตัวใดเป็นตัวลวง

1.2 เนื่องจากการสอบวิเคราะห์ข้อสอบโดยใช้ CDM Package นั้น โปรแกรมจะระบุด้วยว่า ผู้สอบแต่ละคนมีโอกาสในการสอบผ่านคุณลักษณะหนึ่ง ๆ ที่ปรากฏในแบบสอบอยู่เท่าใด ซึ่งจะเป็นประโยชน์



ในการชี้ระบุข้อสอบคุณลักษณะต่างๆ ที่ผู้สอบสอบไม่ผ่านเป็นรายบุคคลได้ ในอนาคต ผู้วิเคราะห์ควรจะหาทางใช้ประโยชน์จากผลการวิเคราะห์ข้อสอบในส่วนนี้ด้วย

2. การวิจัยต่อไป เนื่องจากการพัฒนาแบบสอบวินิจฉัยโดยวิธีการทางพุทธิปัญญายังคงเป็นแนวการพัฒนาแบบสอบที่เพิ่งมีการพัฒนาขึ้น ยังคงมีสาระ ประเด็น ข้อคำถามต่างๆ อีกมากที่ต้องการทำการวิจัยเพื่อหาคำตอบที่ชัดเจน โดยมีประเด็นต่าง ๆ ที่ควรทำวิจัยต่อไปดังนี้

2.1 การวิจัยด้านการพัฒนาแบบสอบ

การพัฒนาแบบสอบวินิจฉัยโดยวิธีการทางพุทธิปัญญาที่ผ่านมายังคงเป็นลักษณะของการทำย้อนหลัง ในความหมายที่ว่า นำเอาแบบสอบที่มีอยู่แล้วมากำหนดคุณลักษณะที่มุ่งศึกษาขึ้นเป็น Q-Matrix แล้วทำการวิเคราะห์ข้อสอบ ลักษณะที่ควรทำในอนาคตคือ การสร้าง Q-Matrix ขึ้นมาก่อนให้มีลักษณะตามที่ควรจะเป็นอย่างแท้จริง สร้างข้อคำถามให้ได้ตามที่ปรากฏใน Q-Matrix ที่มีการวางแผนล่วงหน้าอย่างดี มีการสร้างข้อสอบให้แต่ละข้อวัดคุณลักษณะได้หลายคุณลักษณะ กำหนดให้คุณลักษณะที่ต้องใช้มีความครบถ้วน ครอบคลุม และจัดความเป็นได้ในการเดาตอบถูกออกไปให้มากที่สุด แล้วจึงเก็บข้อมูล และวิเคราะห์ผล

2.2 การเลือกใช้โมเดลในการวิเคราะห์

การวิเคราะห์ข้อสอบที่สร้างขึ้นตามแนวการพัฒนาแบบสอบวินิจฉัยโดยวิธีการทางพุทธิปัญญานั้น มีโมเดลที่ใช้เป็นแนวคิดในการสร้างและวิเคราะห์ข้อสอบหลายโมเดลมาก ผู้พัฒนาแบบสอบควรจะต้องศึกษาโมเดลต่างๆ ให้แตกฉาน รู้ถึงข้อดี ข้อจำกัดของแต่ละโมเดล และควรจะต้องล่วงรู้ว่า แบบสอบที่ตนพัฒนาขึ้นมา นั้น มีลักษณะเหมาะสมที่จะวิเคราะห์โดยใช้โมเดลใด เพื่อให้การใช้โมเดลเป็นไปอย่างถูกต้อง และได้ผลการวิเคราะห์ที่ดีที่สุด

2.3 การตรวจสอบตัวประกอบ

จากผลการวิจัยครั้งนี้ที่พบว่า ค่าประจำข้อสอบโดยวิธีวิเคราะห์ต่างกัน เมื่อนำมาหาตัวประกอบ ได้ตัวประกอบ 2 ตัวหลัก คือ ตัวประกอบด้านค่าความยากที่พ่วงค่าการเดา และตัวประกอบด้านค่าอำนาจจำแนกนั้น ควรมีการทำวิจัยในลักษณะดังกล่าวนี้อีกหลายเรื่อง ในวิชาและกลุ่มประชากรที่แตกต่างกัน เพื่อตรวจสอบความคงทนของตัวประกอบที่ได้ว่า เป็นความจริงที่เป็นทั่วไปเช่นนั้น หรือเป็นความจริงเฉพาะที่เกิดขึ้นกับผลการวิจัยครั้งนี้เท่านั้น โดยเฉพาะหากเป็นไปได้ ควรจะต้องหามาตรการสร้างข้อสอบที่ทำให้ค่าการเดาลดลงมากๆ จนเป็นหรือเกือบเป็นศูนย์ แล้วดูตัวประกอบของค่าประจำข้อสอบที่วิเคราะห์ได้

2.4 ควรจะต้องมีการศึกษาวิจัยว่า จากการที่โปรแกรมวิเคราะห์ข้อสอบในแนวการพัฒนาแบบสอบวินิจฉัยโดยวิธีการทางพุทธิปัญญาสามารถระบุได้ว่า ผู้สอบแต่ละคนมีระดับโอกาสความน่าจะเป็นในการเป็นผู้รอบรู้ในคุณลักษณะต่างๆ แต่ละด้านอยู่เท่าใดนั้น ผลการระบุดังกล่าว มีความเที่ยงตรง (Validity) กับความเป็นจริงของผู้สอบแต่ละคนในระดับใด เพราะส่วนนี้หากผลที่ได้มีความเที่ยงตรงพอ ก็จะเป็นส่วนของงานที่มีประโยชน์ต่อการจัดการเรียนการสอนด้านการแก้ไขพัฒนาผู้เรียนรายบุคคลเป็นอย่างมากในอนาคต



เอกสารอ้างอิง

- เกื้อ กระแสโสม, พิศิษฐ์ ตัณฑวณิช, พนา จินดาศรี, ตรีคม พรหมมาบุญ, บุญทวี อิมบุญตา (2560). การพัฒนาแบบสอบปลายภาครายวิชาการวิจัยเพื่อพัฒนาการเรียนรู้. สุรินทร์ : โปรแกรมทดสอบและวิจัยการศึกษา คณะครุศาสตร์ มหาวิทยาลัยราชภัฏสุรินทร์.
- Assessment Systems Corporation. (1967). **User’s Manual for The XCALIBRE Marginal Maximum-Likelihood Estimation Program.** 2nd. ed. Minnesota.
- George, A.C. ; Robitzsch, A.; Keifer, T.; Groß, J. & Üllü, A. (2016). The R Package CDM for Cognitive Diagnosis Models. **Journal of Statistical Software.**”74(2), 1–4. [Online], Available from : <https://www.jstatsoft.org/article/view/v074i2/0>.Retrived on March25, 2017.
- Lee, Y.S.; de la Torre, J & Park, Y.S. (2012). Relationships between cognitive diagnosis, CTT, and IRT indices : an empirical investigation. **Asia Pacific Education Review.** 13(2): 333 – 345. [Online], Available from [“http://www.researchgate.net/publication/257788305_Relation_between_Cognitive_diagnosis_CTT_IRT_Indices_An_empirical_investigation.”](http://www.researchgate.net/publication/257788305_Relation_between_Cognitive_diagnosis_CTT_IRT_Indices_An_empirical_investigation) Retrieved on April,2 2017.
- Ravand, H. & Robitzsch, A. (2015). Cognitive diagnostic modeling using R. **Practical Assessment, Research & Evaluation.** 20,1-12. สืบค้น 26 March 2016, [Online], Available from : <https://pareonline.net/getvn.asp?v=20&n=11>. Retrieved on March 26, 2016.
- Robitzsch, A., Keifer, T., George, A. C. & Uenlue, A. (2016). **Package CDM** Version 4.8 [Online], Available from : “[https://cran.r-project.org/Package = CDM](https://cran.r-project.org/Package%3DCDM). Retrieved on March 16, 2016.””
- Tatsuoka, K. K. (1984). **Analysis of errors in fraction addition and subtraction problems. Final Report.** University of Illinois at Urbana. [Online], Available from : <https://ericed.gov/?id=ED257665> Retrieved on March 29, 2017.
- Tatsuoka, K. K. (1985). A probabilistic model for diagnosing misconceptions by the pattern classification approach. **Journal of Educational Statistics.** 10, 55–73.
- Tatsuoka, K. K. and Tatsuoka, M. M. (1992). **A psychometrically sound cognitive diagnostic model : effect of remedial as empirical validity.** Educational Testing Service, Princeton New Jersey. [Online], Available from : <https://dtic.mil/tr/fulltest/u2/a255926.pdf> Retrieved on December 27, 2014.