



การเปรียบเทียบคุณภาพของแบบสอบเลือกตอบสองระดับด้านสมรรถนะทางวิทยาศาสตร์
ที่มีระดับความมั่นใจที่แตกต่างกัน โดยประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบแบบตรวจ
ให้คะแนนความรู้บางส่วนแบบทั่วไป



ธัญชา เหล็กทั้ง

การค้นคว้าอิสระเสนอบัณฑิตวิทยาลัย มหาวิทยาลัยนเรศวร
เพื่อเป็นส่วนหนึ่งของการศึกษา หลักสูตรการศึกษามหาบัณฑิต
สาขาวิชานวัตกรรมทางการวัดผลการเรียนรู้

ปีการศึกษา 2566

ลิขสิทธิ์เป็นของมหาวิทยาลัยนเรศวร

การเปรียบเทียบคุณภาพของแบบสอบเลือกตอบสองระดับด้านสมรรถนะทางวิทยาศาสตร์
ที่มีระดับความมั่นใจที่แตกต่างกัน โดยประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบแบบตรวจ
ให้คะแนนความรู้บางส่วนแบบทั่วไป



การค้นคว้าอิสระเสนอบัณฑิตวิทยาลัย มหาวิทยาลัยนเรศวร
เพื่อเป็นส่วนหนึ่งของการศึกษา หลักสูตรการศึกษามหาบัณฑิต
สาขาวิชานวัตกรรมทางการวัดผลการเรียนรู้
ปีการศึกษา 2566
ลิขสิทธิ์เป็นของมหาวิทยาลัยนเรศวร

การค้นคว้าอิสระ เรื่อง "การเปรียบเทียบคุณภาพของแบบสอบเลือกตอบสองระดับด้านสมรรถนะทาง
วิทยาศาสตร์ที่มีระดับความมั่นใจที่แตกต่างกัน โดยประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบแบบตรวจ
ให้คะแนนความรู้บางส่วนแบบทั่วไป"

ของ ธนัชชา เหล็กทั้ง

ได้รับการพิจารณาให้นับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาการศึกษามหาบัณฑิต สาขาวิชานวัตกรรมการวัดผลการเรียนรู้

.....
(รองศาสตราจารย์ ดร.น้ำทิพย์ อองอาจวานิชย์)

อาจารย์ที่ปรึกษาการค้นคว้าอิสระ

.....
(ผู้ช่วยศาสตราจารย์ ดร.อัจฉรา ศรีพันธ์)

หัวหน้าภาควิชาบริหาร วิจัย และพัฒนานวัตกรรมการศึกษา



ชื่อเรื่อง	การเปรียบเทียบคุณภาพของแบบสอบเลือกตอบสองระดับด้านสมรรถนะทางวิทยาศาสตร์ที่มีระดับความมั่นใจที่แตกต่างกัน โดยประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบแบบตรวจให้คะแนนความรู้บางส่วนแบบทั่วไป
ผู้วิจัย	ธัญชา เหล็กทัง
ประธานที่ปรึกษา	รองศาสตราจารย์ ดร.น้ำทิพย์ ่องอาจวานิชย์
ประเภทสารนิพนธ์	การค้นคว้าอิสระ กศ.ม. นวัตกรรมทางการวัดผลการเรียนรู้, มหาวิทยาลัยนเรศวร, 2566
คำสำคัญ	ระดับความมั่นใจ, ทฤษฎีการตอบสนองข้อสอบ, โมเดลการตอบสนองข้อสอบแบบตรวจให้คะแนนความรู้บางส่วนแบบทั่วไป

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบ 1) ความยากและอำนาจจำแนก 2) ความตรงเชิงโครงสร้าง 3) ความเที่ยง และ 4) ฟังก์ชันสารสนเทศของแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่แตกต่างกัน กลุ่มตัวอย่างในการวิจัยได้จากการสุ่มแบบหลายขั้นตอน จากนักเรียนระดับชั้นมัธยมศึกษาปีที่ 4-6 สังกัดสำนักงานเขตพื้นที่การศึกษามัธยมศึกษาพิษณุโลก อุตรดิตถ์ จำนวน 39 โรงเรียนในเขตจังหวัดพิษณุโลก ได้กลุ่มตัวอย่างจำนวน 750 คน แบ่งเป็นกลุ่มตัวอย่าง 2 กลุ่ม กลุ่มละ 375 คน เครื่องมือที่ใช้ในการวิจัย ได้แก่ แบบสอบเลือกตอบด้านสมรรถนะทางวิทยาศาสตร์ จำนวน 2 ฉบับ ฉบับละ 20 ข้อ โดยฉบับที่ 1 มีระดับความมั่นใจ 2 ระดับ ได้แก่ มั่นใจ และไม่มั่นใจ และฉบับที่ 2 มีระดับความมั่นใจ 4 ระดับ ได้แก่ มั่นใจมาก มั่นใจ ไม่มั่นใจ และไม่มั่นใจ และ เดา

ผลการวิจัยพบว่า แบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 2 ระดับ มีคุณภาพดีกว่าแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 4 ระดับ โดยพิจารณาองค์ประกอบดังนี้ 1) การตรวจสอบข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ พบว่า แบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 4 ระดับ ไม่ผ่านเกณฑ์ด้านอัตราส่วนของค่าไอเกน แต่แบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 2 ระดับ ผ่านเกณฑ์การพิจารณาทั้ง 2 เกณฑ์ 2) ความยากและอำนาจจำแนกเฉลี่ยของแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 2 ระดับ มีค่าความยากเฉลี่ยที่

เหมาะสม รวมทั้งยังสามารถจำแนกผู้สอบได้ดีกว่าแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 4 ระดับ 3) ความตรงเชิงโครงสร้างของแบบสอบทั้ง 2 ฉบับ มีความตรงเชิงโครงสร้าง แต่เมื่อพิจารณาน้ำหนักองค์ประกอบของแบบสอบแต่ละฉบับ พบว่า แบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 4 ระดับ มีข้อสอบที่ไม่ผ่านเกณฑ์น้ำหนักองค์ประกอบมากกว่าแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 2 ระดับ 4) ความเที่ยงของแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 2 ระดับ มีค่ามากกว่าแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 4 ระดับ และ 5) ฟังก์ชันสารสนเทศของแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 2 ระดับ สามารถสะท้อนความสามารถที่แท้จริงของผู้สอบได้มากกว่าแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 4 ระดับ



Title	COMPARISON OF THE QUALITY OF TWO-LEVEL MULTIPLE-CHOICE TESTS IN SCIENTIFIC COMPETENCY WITH DIFFERENT CONFIDENCE LEVELS, APPLYING THE GENERALIZED PARTIAL CREDIT MODEL FROM ITEM RESPONSE THEORY
Author	Thanatcha Lekthang
Advisor	Associate Professor Namthip Ongardwanich, Ph.D.
Academic Paper	M.Ed. Independent Study in Innovation of Learning Measurement - (Plan B) (Teacher), Naresuan University, 2023
Keywords	confidence level, item response theory, generalized partial credit model

ABSTRACT

The purposes of this research aim to compare: 1) difficulty and discriminant index, 2) construct validity, 3) reliability coefficient, and 4) test information of tests derived from the partial knowledge scoring method when tests have different confidence levels. The sample was obtained through multi-stage sampling from high school students in grades 10-12 under the Secondary Educational Service Area Office Phitsanulok Uttaradit, totaling 39 schools in Phitsanulok province. The sample consisted of 750 students, divided into 2 groups of 375 each. Research instruments were 2 versions of multiple-choice tests on scientific competency, each with 20 items. The first version had 2 confidence levels: confident and not confident. The second version had 4 confidence levels: very confident, confident, not confident and guessing.

The results showed that tests derived from the partial knowledge scoring method with 2 confidence levels were of higher quality than those with 4 confidence levels. The considerations are as follows: 1) examining the basic assumptions of the item response theory indicated that tests with 4 confidence levels did not meet the eigenvalue ratio criterion, while tests with 2 confidence levels meet both criteria. 2)

the average difficulty and discrimination index of tests with 2 confidence levels were appropriate and could better discriminate examinees than those with 4 confidence levels. 3) both versions of the tests exhibited construct validity, but when considering the factors loading of each test, the tests with 4 confidence levels had more items failing the factor loading criterion than the tests with 2 confidence levels. 4) the reliability of tests with 2 confidence levels was higher than that of tests with 4 confidence levels and 5) the test information of tests with 2 confidence levels reflected the true ability of examinees more accurately than tests with 4 confidence levels.



ประกาศคุณูปการ

การค้นคว้าอิสระฉบับนี้ สำเร็จลงได้ด้วยความรู้ความกรุณาอย่างยิ่งจากอาจารย์ที่ปรึกษา รองศาสตราจารย์ ดร.น้ำทิพย์ อองอาจวานิชย์ ที่ได้สละเวลาอันมีค่าในการให้คำแนะนำตลอดระยะเวลาในการทำการค้นคว้าอิสระฉบับนี้ ตลอดจนการแก้ไขข้อบกพร่อง นอกจากนี้ขอขอบพระคุณอาจารย์ประจำสาขาวิศวกรรมทางการวัดผลการเรียนรู้ คณะศึกษาศาสตร์ มหาวิทยาลัยรัตนนครทุกท่าน ที่ได้มอบความรู้ ประสบการณ์ และการดูแลเอาใจใส่อย่างดียิ่ง จนการค้นคว้าอิสระฉบับนี้สำเร็จสมบูรณ์ได้ ผู้วิจัยขอขอบพระคุณไว้ ณ ที่นี้

ขอขอบพระคุณ รองศาสตราจารย์ ดร.กฤษฎาภาณุจัน โทพิทักษ์ ผู้ช่วยศาสตราจารย์ ดร.ชำนาญ ปาณาวงษ์ นางกุลกรวี ขาวสง่า ผู้ทรงคุณวุฒิในการตรวจเครื่องมือการค้นคว้าอิสระในครั้งนี้ ที่ได้กรุณาให้คำแนะนำ ตลอดจนแก้ไขตรวจสอบเครื่องมือที่ใช้ในการวิจัยด้วยความเอาใจใส่ จนทำให้การค้นคว้าอิสระฉบับนี้สำเร็จลุล่วงได้อย่างสมบูรณ์และมีคุณค่า

ขอขอบพระคุณผู้บริหารสถานศึกษา และคณะครูกลุ่มสาระการเรียนรู้วิทยาศาสตร์และเทคโนโลยีของโรงเรียนกลุ่มตัวอย่างทั้ง 39 โรงเรียน ที่ให้ความอนุเคราะห์ อำนวยความสะดวก และประสานงานกับนักเรียนในการเก็บข้อมูลเพื่อการค้นคว้าอิสระในครั้งนี้

ขอขอบคุณนักเรียนของโรงเรียนกลุ่มตัวอย่างทั้ง 750 คน ที่ตั้งใจ และให้ความร่วมมือในการเก็บรวบรวมข้อมูล

เหนือสิ่งอื่นใดขอขอบพระคุณ คุณพ่อ คุณแม่ และครอบครัวของผู้วิจัย ที่คอยให้กำลังใจ และให้การสนับสนุนในทุก ๆ ด้านอย่างดีที่สุดเสมอมา

คุณค่าและคุณประโยชน์อันพึงจะมีจากการค้นคว้าอิสระฉบับนี้ ผู้วิจัยขอมอบและอุทิศแด่ผู้มีพระคุณทุก ๆ ท่าน

ธัญชา เหล็กทั้ง

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ค
บทคัดย่อภาษาอังกฤษ.....	จ
ประกาศคุุณุปการ.....	ช
สารบัญ.....	ซ
สารบัญตาราง.....	ญ
สารบัญภาพ.....	ฉ
บทที่ 1.....	1
ความเป็นมาของปัญหา.....	1
คำถามของการวิจัย.....	4
วัตถุประสงค์ของการวิจัย.....	5
ขอบเขตของการวิจัย.....	5
นิยามศัพท์เฉพาะ.....	6
ประโยชน์ที่ได้รับจากการวิจัย.....	9
บทที่ 2.....	10
ตอนที่ 1 แนวคิดเกี่ยวกับแบบสอบถามเลือกตอบ.....	10
ตอนที่ 2 แนวคิดเกี่ยวกับทฤษฎีการตอบสนองข้อสอบ.....	20
ตอนที่ 3 แนวคิดเกี่ยวกับวิธีการตรวจให้คะแนนความรู้บางส่วน.....	29
ตอนที่ 4 แนวคิดเกี่ยวกับสมรรถนะทางวิทยาศาสตร์.....	35
ตอนที่ 5 กรอบแนวคิดการวิจัย.....	41

บทที่ 3	43
ประชากรและกลุ่มตัวอย่าง.....	43
เครื่องมือที่ใช้ในการวิจัย.....	47
การเก็บรวบรวมข้อมูล.....	54
การวิเคราะห์ข้อมูล	56
บทที่ 4	61
ตอนที่ 1 ผลการวิเคราะห์ความยากและอำนาจจำแนกของแบบสอบ.....	61
ตอนที่ 2 ผลการวิเคราะห์ความตรงเชิงโครงสร้างของแบบสอบ.....	65
ตอนที่ 3 ผลการวิเคราะห์ความเที่ยงของแบบสอบ	74
ตอนที่ 4 ผลการวิเคราะห์ฟังก์ชันสารสนเทศของแบบสอบ	74
ตอนที่ 5 ผลการเปรียบเทียบคุณภาพของแบบสอบ.....	76
บทที่ 5	78
สรุปผลการวิจัย.....	79
อภิปรายผลการวิจัย.....	80
ข้อเสนอแนะในการนำงานวิจัยไปใช้.....	83
ข้อเสนอแนะในการวิจัยครั้งต่อไป	84
บรรณานุกรม	85
ภาคผนวก.....	89
ประวัติผู้วิจัย	123

สารบัญตาราง

	หน้า
ตาราง 1 ลักษณะของแบบสอบแบบเลือกตอบสองระดับ สามระดับ และสี่ระดับ.....	19
ตาราง 2 ข้อดีและข้อจำกัดของแบบสอบแบบเลือกตอบสองระดับ สามระดับ และสี่ระดับ.....	19
ตาราง 3 สรุปลักษณะของโมเดลการตอบสนองข้อสอบที่มีการตรวจให้คะแนนมากกว่าสองค่า.....	29
ตาราง 4 บริบทที่ใช้ในการประเมิน 3 ระดับ กับตัวอย่างสถานการณ์ที่เป็นส่วนหนึ่งในชีวิตประจำวัน.....	37
ตาราง 5 ประชากรจำแนกตามจังหวัด จำนวนโรงเรียน จำนวนห้องเรียน และจำนวนนักเรียน.....	43
ตาราง 6 จำนวนกลุ่มตัวอย่างจำแนกตามอำเภอในจังหวัดพิษณุโลก.....	44
ตาราง 7 จำนวนกลุ่มตัวอย่างจำแนกตามโรงเรียนในจังหวัดพิษณุโลก.....	45
ตาราง 8 องค์ประกอบและพฤติกรรมบ่งชี้ของสมรรถนะทางวิทยาศาสตร์.....	48
ตาราง 9 แผนผังการสร้างแบบสอบเลือกตอบด้านสมรรถนะทางวิทยาศาสตร์.....	50
ตาราง 10 ผลการวิเคราะห์ค่าความยากและอำนาจจำแนกของแบบสอบเลือกตอบด้านสมรรถนะทางวิทยาศาสตร์.....	53
ตาราง 11 เกณฑ์การพิจารณาดัชนีความสอดคล้องระหว่างโมเดลและข้อมูล.....	59
ตาราง 12 ผลการตรวจสอบข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ.....	62
ตาราง 13 ผลการวิเคราะห์ความยากและอำนาจจำแนกของแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่แตกต่างกัน.....	63
ตาราง 14 ผลการวิเคราะห์ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของความยากและอำนาจจำแนกของแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่แตกต่างกัน.....	64

ตาราง 15 ผลการวิเคราะห์ความแตกต่างของค่าเฉลี่ยของค่าความยากเฉลี่ยของแบบสอบ ที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่แตกต่าง กัน	64
ตาราง 16 ผลการวิเคราะห์ความแตกต่างของค่าเฉลี่ยของค่าอำนาจจำแนกเฉลี่ยของแบบ สอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่ แตกต่างกัน.....	65
ตาราง 17 ผลการวิเคราะห์ความสอดคล้องระหว่างโมเดลกับข้อมูลที่ได้จากวิธีการตรวจให้ คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่แตกต่างกัน	68
ตาราง 18 ผลการวิเคราะห์น้ำหนักองค์ประกอบของโมเดลที่ได้จากวิธีการตรวจให้คะแนน ความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 2 ระดับ	69
ตาราง 19 ผลการวิเคราะห์น้ำหนักองค์ประกอบของโมเดลที่ได้จากวิธีการตรวจให้คะแนน ความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 4 ระดับ	72
ตาราง 20 ผลการวิเคราะห์ความเที่ยงของแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้ บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่แตกต่างกัน	74
ตาราง 21 ผลการวิเคราะห์ฟังก์ชันสารสนเทศของแบบสอบที่ได้จากวิธีการตรวจให้คะแนน ความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่แตกต่างกัน	75
ตาราง 22 เปรียบเทียบคุณภาพของแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้ บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่แตกต่างกัน	77

สารบัญภาพ

	หน้า
ภาพ 1 โมเดลอธิบายการเดาตามแนวคิดของ.....	13
ภาพ 2 ปัจจัยที่ส่งผลต่อการเดาในแบบสอบถามเลือกตอบ.....	16
ภาพ 3 โค้งลักษณะข้อสอบ.....	21
ภาพ 4 โค้งรายการคำตอบ (CRC) จากการวิเคราะห์โดยใช้ G-PCM.....	27
ภาพ 5 องค์ประกอบของกรอบการประเมินด้านวิทยาศาสตร์.....	36
ภาพ 6 กรอบแนวคิดในการวิจัย.....	42
ภาพ 7 โครงสร้างของแบบสอบสำหรับการวิเคราะห์องค์ประกอบเชิงยืนยัน.....	58
ภาพ 8 โครงสร้างของแบบสอบสำหรับการวิเคราะห์องค์ประกอบเชิงยืนยัน.....	67
ภาพ 9 โมเดลที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความ มั่นใจ 2 ระดับ.....	71
ภาพ 10 โมเดลที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความ มั่นใจ 4 ระดับ.....	73
ภาพ 11 โค้งฟังก์ชันสารสนเทศของแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้ บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 2 ระดับ.....	75
ภาพ 12 โค้งฟังก์ชันสารสนเทศของแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้ บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 4 ระดับ.....	76

บทที่ 1

บทนำ

ความเป็นมาของปัญหา

การเปลี่ยนแปลงในโลกยุคปัจจุบัน รวมถึงสภาพการแข่งขันอย่างรุนแรง ทำให้องค์กรทุกแห่งต้องเผชิญกับความเสี่ยงที่สูงขึ้น อุตสาหกรรมหลายแห่งกำลังเผชิญภาวะการพลิกผันทางดิจิทัล (Digital Disruption) ที่ทำให้วัฏกรรมทางเทคโนโลยีใหม่ ๆ เปลี่ยนแปลงรูปแบบวิธีการ รวมถึงพฤติกรรมของคนในสังคมไปอย่างฉับพลัน ทั้งทางธุรกิจ สังคม และวัฒนธรรม ทำให้การพัฒนาทรัพยากรมนุษย์ในปัจจุบันต้องการพัฒนามนุษย์ที่อุดมไปด้วยความรู้ ทักษะ และความสามารถที่พร้อมจะต่อสู้กับการเปลี่ยนแปลง หรือที่เรียกว่า สมรรถนะ (Competency) ซึ่งเป็นคุณสมบัติที่ช่วยให้บุคคลประสบความสำเร็จในการทำงาน สมรรถนะจึงเป็นผลลัพธ์ที่พึงประสงค์ของการศึกษาและการเรียนรู้ ดังนั้นระบบการศึกษาในปัจจุบันจึงควรช่วยกันพัฒนาผู้เรียนให้เกิดสมรรถนะสำคัญที่จะเป็นประโยชน์ต่อการดำรงชีวิตอย่างมีคุณภาพ (สำนักวิชาการและมาตรฐานการศึกษา, 2564) การมีเพียงความรู้และทักษะนั้นยังไม่ถือว่าเป็นสมรรถนะ จนกว่าบุคคลนั้นจะสามารถนำความรู้และทักษะมาประยุกต์ใช้กับงาน ทำงานให้เกิดผลลัพธ์ที่ได้อย่างชัดเจน ประสบความสำเร็จในการปฏิบัติ จึงจะถือว่าบุคคลนั้นมีสมรรถนะ

การแก้ไขปัญหาก็เกี่ยวข้องกับบริบททางวิทยาศาสตร์ อาศัยการนำเอาความรู้ด้านเนื้อหาทางวิทยาศาสตร์ผนวกเข้ากับสมรรถนะทางวิทยาศาสตร์ ซึ่งสามารถแบ่งออกเป็น 3 ด้าน ได้แก่ 1) การอธิบายปรากฏการณ์ในเชิงวิทยาศาสตร์ (Explain Phenomena Scientifically) คือ ความสามารถของบุคคลในการใช้ความรู้ด้านเนื้อหาเพื่อแปลความและให้คำอธิบายกับปรากฏการณ์ทั้งที่เกิดขึ้นในชีวิตประจำวัน หรืออาจเป็นปรากฏการณ์ในบริบทที่ไม่คุ้นเคยได้อย่างเหมาะสม 2) การประเมินและออกแบบกระบวนการสืบเสาะหาความรู้ทางวิทยาศาสตร์ (Evaluate and Design Scientific Enquiry) คือ ความสามารถของบุคคลในการประเมินข้อค้นพบทางวิทยาศาสตร์อย่างมีวิจารณญาณ สามารถแยกแยะได้ว่าคำถามหรือปัญหาใดสามารถตอบหรือแก้ไขได้ด้วยวิธีการทางวิทยาศาสตร์ รวมถึงการตระหนักถึงความสำคัญและคุณค่าของงานวิจัยที่ส่งผลต่อการค้นคว้าทางวิทยาศาสตร์ในอนาคต และความสำคัญของการตั้งข้อสงสัยในการรายงานการวิจัยทางวิทยาศาสตร์ และ 3) การแปลความหมายข้อมูลและการใช้ประจักษ์พยานในเชิงวิทยาศาสตร์ (Interpret Data and Evidence Scientifically) คือ ความสามารถของบุคคลในการตีความข้อมูลและหลักฐานทางวิทยาศาสตร์ที่ใช้ในการสร้างคำกล่าวอ้างหรือลงข้อสรุป และนำเสนอข้อมูลที่ได้รับมาในรูปแบบอื่น นอกจากนี้จะต้อง

สามารถสร้างข้อสรุปที่สมเหตุสมผลบนพื้นฐานของประจักษ์พยาน ข้อมูล หรือประเมินข้อสรุปของผู้อื่นได้ว่าสอดคล้องกับประจักษ์พยานที่มีหรือไม่มี รวมถึงสามารถให้เหตุผลสนับสนุนหรือโต้แย้งข้อสรุปเหล่านั้นได้อย่างสมเหตุสมผล (สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี, 2564) การที่บุคคลจะมีสมรรถนะทางวิทยาศาสตร์ที่จำเป็นต่อการรู้วิทยาศาสตร์ได้ต้องอาศัยความรู้ทางวิทยาศาสตร์ด้านเนื้อหา (Content Knowledge) เป็นความรู้ที่เกี่ยวข้องกับธรรมชาติที่แสดงประสบการณ์ตั้งแต่ระดับบุคคล ท้องถิ่น ประเทศ และระดับโลก ด้านกระบวนการ (Procedural Knowledge) เป็นความรู้ที่นักวิทยาศาสตร์นำมาใช้เพื่อให้ได้มาซึ่งข้อมูลที่น่าเชื่อถือ และสมเหตุสมผล ความรู้ด้านนี้จำเป็นต่อการสืบเสาะทางวิทยาศาสตร์ และการพิจารณาหลักฐานที่จะนำมาใช้สนับสนุนข้อสรุปทางวิทยาศาสตร์ และด้านการได้มาของความรู้ (Epistemic Knowledge) ตามกรอบแนวทางของ PISA ได้นำเสนอไว้ 2 ประเด็นคือ แนวคิดและการให้นิยามเกี่ยวกับลักษณะสำคัญของวิทยาศาสตร์ และบทบาทของแนวคิดและลักษณะสำคัญที่ใช้ในการประเมินความรู้ที่ได้จากวิทยาศาสตร์ (สำนักวิชาการและมาตรฐานการศึกษา, 2558)

ผลการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) วิชาวิทยาศาสตร์ และคณิตศาสตร์ พบว่า คะแนนเฉลี่ยของนักเรียนไทยยังต่ำกว่าเกณฑ์มาตรฐาน และผลการประเมินสมรรถนะนักเรียนมาตรฐานสากล (Program for International Student Assessment หรือ PISA) รอบปี 2018 จาก 79 ประเทศทั่วโลกพบว่า คะแนนวิทยาศาสตร์ และคณิตศาสตร์ของประเทศไทยมีแนวโน้มคงที่เมื่อเทียบกับปีก่อน ๆ แต่คะแนนการอ่านลดลง ซึ่งสะท้อนให้เห็นว่าความฉลาดรู้ (Literacy) หรือสมรรถนะในการเชื่อมโยงความรู้แนวคิดวิทยาศาสตร์ และคณิตศาสตร์เพื่อการวิเคราะห์สถานการณ์หรือปัญหา การหาคำตอบ แสดงเหตุผล หรือลงข้อสรุปได้อย่างมีวิจารณญาณสามารถสื่อสารได้อย่างเหมาะสมของนักเรียนไทยยังอยู่ในระดับที่ต่ำกว่าค่าเฉลี่ย ประเทศไทยจึงจำเป็นต้องยกระดับคุณภาพทางการศึกษา โดยปรับเปลี่ยนการเรียนรู้จากเดิมที่เคยเน้นเนื้อหาวิชาเป็นการสอนเพื่อพัฒนาสมรรถนะและความฉลาดรู้ของนักเรียนมากกว่าที่จะเน้นให้เกิดการพัฒนาความรู้แต่เพียงอย่างเดียว เพราะความฉลาดรู้และสมรรถนะเป็นสิ่งจำเป็นสำหรับคนไทยที่ต้องมีเพื่อเป็นทักษะพื้นฐานในศตวรรษที่ 21 (สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี, 2564) สิ่งจำเป็นเร่งด่วนที่ต้องเปลี่ยนแปลงคือ การพัฒนาครูให้สามารถจัดการเรียนการสอนฐานสมรรถนะ การพัฒนารูปแบบการวัดและประเมินผลการเรียนรู้ที่วัดการนำไปประยุกต์ใช้ การคิดวิเคราะห์ และการแก้ปัญหามากกว่าวัดความจำ หรือที่เรียกว่า การวัดผลเชิงสมรรถนะ (ชูกิจ ลิมปิจำนงค์, 2564)

ลักษณะของเครื่องมือที่เคยใช้ในการประเมินสมรรถนะทางวิทยาศาสตร์ในการสอบ PISA 2006 นั้นมีข้อสอบหลายลักษณะ เช่น การให้นักเรียนเลือกข้อสรุปที่สมเหตุสมผลกับประจักษ์พยานหรือข้อมูลจากตัวเลือกที่มีให้ การให้นักเรียนบอกเหตุผลสนับสนุนหรือคัดค้านการลงข้อสรุปที่ได้มาจากกระบวนการหนึ่ง ๆ การให้นักเรียนอธิบายหรือให้เหตุผลว่า กระบวนการนั้นควรนำมาสู่ข้อสรุป

นั้นหรือไม่อย่างไร การให้คิดแบบวิพากษ์วิจารณ์ว่า ข้อตกลงเบื้องต้น (Assumption) ที่กำหนดไว้สำหรับการศึกษาแล้วนำไปสู่ข้อสรุปนั้นเหมาะสมหรือไม่อย่างไร (สุนีย์ คล้ายนิล และอัมพิกา ประโมจน์ย์, 2551: 22) ในการวัดและประเมินผลการศึกษาเป็นองค์ประกอบสำคัญในการพัฒนาคุณภาพทางการศึกษา เพื่อให้ได้ข้อมูลสารสนเทศที่จำเป็นในการพิจารณาระดับคุณภาพการเรียนรู้ของผู้เรียน ซึ่งแผนการศึกษาแห่งชาติ พ.ศ.2560-2579 ยุทธศาสตร์ที่ 3 การพัฒนาศักยภาพคนทุกช่วงวัย และการสร้างสังคมแห่งการเรียนรู้ เป้าหมายที่ 5 ว่าด้วยการมีระบบและกลไกการทดสอบ การวัดและประเมินความรู้ ทักษะ และสมรรถนะของผู้เรียนทุกระดับการศึกษา และทุกกลุ่มเป้าหมายที่มีประสิทธิภาพ (สำนักงานเลขาธิการสภาการศึกษา กระทรวงศึกษาธิการ, 2560) โดยแบบสอบแบบหลายตัวเลือก (Multiple Choice) เป็นแบบสอบที่นิยมใช้กันอย่างแพร่หลาย มีความสะดวกในการนำไปใช้หลายประการ เช่น วัดพฤติกรรมการเรียนรู้ได้หลากหลายและครอบคลุมเนื้อหา มีความเป็นปรนัยในการตรวจให้คะแนน ใช้เวลาในการตรวจน้อย สามารถนำผลการตอบไปวิเคราะห์ข้อมูลทางสถิติได้ เหมาะสำหรับการสอบที่มีผู้สอบจำนวนมาก เป็นต้น (โชติกา ภาชีผล, ณีภูธรณ์ หลาวทอง และกมลวรรณ ตังธนากานนท์, 2558) แต่แบบสอบหลายตัวเลือกยังมีข้อจำกัดหลายด้านที่ต้องพึงระวัง เช่น ผู้สอบมีโอกาสในการเดาคำตอบได้ถูก สามารถลอกคำตอบกันได้ง่าย ไม่เหมาะสำหรับการวัดพฤติกรรมในระดับการสังเคราะห์ อีกทั้งรูปแบบการตรวจให้คะแนนของแบบสอบหลายตัวเลือกในปัจจุบันมีการให้คะแนนแบบประเพณีนิยม นั่นคือ ตอบถูกได้ 1 คะแนน และตอบผิดได้ 0 คะแนน ดังนั้นผู้สอบจึงมีโอกาสในการเดาข้อสอบได้ ซึ่งจะทำได้ไม่สะท้อนความสามารถที่แท้จริงของผู้เรียน และไม่สามารถให้สารสนเทศเพียงพอที่จะจำแนกผู้เรียนที่มีความรู้จริง (Full Knowledge) มีความรู้บางส่วน (Partial Knowledge) หรือไม่มีความรู้ (Absence of Knowledge) (เอมอร์ จังศิริพรปรกรณ์, 2545b) จากข้อจำกัดดังกล่าว จึงมีการศึกษาวิธีการแก้ปัญหาการเดาคำตอบของผู้สอบและการให้สารสนเทศในการจำแนกผู้เรียนโดยใช้วิธีการตรวจให้คะแนนความรู้บางส่วน ที่แสดงถึงตำแหน่งจริงของความรู้ของผู้ตอบ ซึ่งมีส่วนช่วยลดความคลาดเคลื่อนจากแหล่งการเดาของผู้ตอบลง ทำให้คะแนนมีความเที่ยงมากขึ้น ในขณะเดียวกันก็จะส่งผลต่อความตรงของกระบวนการวัดด้วย (Frary, 1980)

การทดสอบแบบเลือกตอบให้มีสองระดับ (Two-tier Multiple Choice Test) จึงเป็นอีกรูปแบบหนึ่งของการทดสอบที่มีความซับซ้อนมากกว่าแบบสอบแบบหลายตัวเลือกทั่ว ๆ ไป โดยลักษณะของแบบสอบเลือกตอบสองระดับ ประกอบด้วย ระดับแรกหรือระดับเนื้อหา (Content Tier) วัดความรู้เนื้อหาสำคัญด้วยแบบสอบเลือกตอบ 4 ตัวเลือก ซึ่งเป็นรูปแบบของแบบสอบเลือกตอบหลายตัวเลือกแบบดั้งเดิม เกี่ยวข้องกับองค์ความรู้เนื้อหาสำคัญ ระดับสองหรือระดับความมั่นใจ (Confidence Tier) วัดความมั่นใจของคำตอบในระดับเนื้อหา พิจารณาคุณลักษณะภายในของแต่ละ

บุคคล ซึ่งในการศึกษาทางวิทยาศาสตร์ยังพบงานวิจัยไม่มากนักที่นำมาตรปรมาณความมั่นใจมาใช้ ในการพัฒนาแบบสอบ

จากการศึกษาทฤษฎีและงานวิจัยที่ผ่านมาพบว่า การศึกษาคุณภาพของแบบสอบจะเป็นการ เปรียบเทียบในด้านความตรงและความเที่ยงเป็นส่วนใหญ่ โดยเป็นการใช้ทฤษฎีการวัดแบบดั้งเดิม (Classical Test Theory) ซึ่งมีข้อจำกัดในเรื่องแหล่งความคลาดเคลื่อนในการวัด และค่าสถิติที่แปร เปลี่ยนไปตามกลุ่มตัวอย่างที่ใช้ ปัจจุบันจึงมีการพัฒนาทฤษฎีการตอบสนองข้อสอบ (Item Response Theory) เพื่อแก้ปัญหาการวิเคราะห์ตามแนวทฤษฎีการวัดแบบดั้งเดิม เป็นการพัฒนา โมเดลและตรวจสอบโมเดล สร้างมาตรวัดคะแนนความสามารถของผู้สอบ การพัฒนาคะแนน ความสามารถในรูปฟังก์ชันของพารามิเตอร์ความสามารถของผู้สอบรูปแบบต่าง ๆ โดยมีการกำหนด คะแนนแบบต่าง ๆ และการพัฒนาวิธีประมาณค่าพารามิเตอร์ความสามารถแบบต่าง ๆ ด้วย

จากปัญหาที่กล่าวมาข้างต้น ทำให้ผู้วิจัยสนใจศึกษาการเปรียบเทียบคุณภาพภาพของวิธีการ ตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่แตกต่างกัน ได้แก่ ระดับความมั่นใจ 2 ระดับ และระดับความมั่นใจ 4 ระดับ ของแบบสอบเลือกตอบสองระดับด้านสมรรถนะทาง วิทยาศาสตร์ โดยประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบแบบตรวจให้คะแนนความรู้บางส่วนแบบ ทั่วไป เพื่อขยายองค์ความรู้เกี่ยวกับวิธีการตรวจให้คะแนนความรู้บางส่วน และเป็นแนวทางในการวัด และประเมินผลการเรียนรู้ของผู้เรียนที่สะท้อนความสามารถที่แท้จริงของผู้เรียนได้

คำถามของการวิจัย

1. ความยากและอำนาจจำแนกของแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่แตกต่างกัน มีความแตกต่างกันหรือไม่ อย่างไร
2. ความตรงเชิงโครงสร้างของแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อ ข้อสอบมีระดับความมั่นใจที่แตกต่างกัน มีความแตกต่างกันหรือไม่ อย่างไร
3. ความเที่ยงของแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมี ระดับความมั่นใจที่แตกต่างกัน มีความแตกต่างกันหรือไม่ อย่างไร
4. ฟังก์ชันสารสนเทศของแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อ ข้อสอบมีระดับความมั่นใจที่แตกต่างกัน มีความแตกต่างกันหรือไม่ อย่างไร

วัตถุประสงค์ของการวิจัย

1. เพื่อเปรียบเทียบความยากและอำนาจจำแนกของแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่แตกต่างกัน
2. เพื่อเปรียบเทียบความตรงเชิงโครงสร้างของแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่แตกต่างกัน
3. เพื่อเปรียบเทียบความเที่ยงของแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่แตกต่างกัน
4. เพื่อเปรียบเทียบฟังก์ชันสารสนเทศของแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่แตกต่างกัน

ขอบเขตของการวิจัย

1. ประชากรที่ใช้ในการวิจัย ได้แก่ นักเรียนระดับชั้นมัธยมศึกษาปีที่ 4-6 จำนวน 21,663 คน จำนวน 57 โรงเรียน สังกัดสำนักงานเขตพื้นที่การศึกษามัธยมศึกษาพิษณุโลก อุตรดิตถ์ (สำนักงานเขตพื้นที่การศึกษามัธยมศึกษาพิษณุโลก อุตรดิตถ์, 2566)

2. ตัวแปรที่ศึกษา

2.1 *ตัวแปรต้น* ได้แก่ วิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่แตกต่างกัน 2 ระดับ ได้แก่

2.1.1 ระดับความมั่นใจ 2 ระดับ

2.1.2 ระดับความมั่นใจ 4 ระดับ

2.2 *ตัวแปรตาม* ได้แก่ คุณภาพของวิธีการตรวจให้คะแนนความรู้บางส่วน 4 ด้าน ได้แก่

2.2.1 ความยากและอำนาจจำแนกของแบบสอบ

2.2.2 ความตรงเชิงโครงสร้างของแบบสอบ

2.2.3 ความเที่ยงของแบบสอบ

2.2.4 ฟังก์ชันสารสนเทศของแบบสอบ

3. *ขอบเขตด้านเนื้อหา* ได้แก่ เนื้อหารายวิชาวิทยาศาสตร์ชีวภาพ จำนวน 5 เรื่อง คือ 1) การลำเลียงสาร 2) การรักษาคุณภาพของร่างกายมนุษย์ 3) การดำรงชีวิตของพืช 4) พันธุกรรมและวิวัฒนาการ และ 5) ชีวิตในสิ่งแวดล้อม ที่มีความสอดคล้องกับสมรรถนะทางวิทยาศาสตร์ที่ต้องการจะวัดทั้ง 3 องค์ประกอบ ดังนี้ *องค์ประกอบที่ 1 การอธิบายปรากฏการณ์ในเชิงวิทยาศาสตร์ (Explain Phenomena Scientifically)* โดยบุคคลต้องสามารถระลึกถึงความรู้ด้านเนื้อหาที่เหมาะสมในสถานการณ์ที่กำหนดให้ และใช้ความรู้เพื่อแปลความหมายและให้คำอธิบายต่อ

ปรากฏการณ์ต่าง ๆ รวมถึงการวาดแบบจำลองทางวิทยาศาสตร์เพื่อใช้อธิบายปรากฏการณ์ที่เกิดขึ้นในชีวิตประจำวัน การบรรยายและการตีความปรากฏการณ์ การคาดการณ์หรือการพยากรณ์การเปลี่ยนแปลงที่อาจจะเกิดขึ้น รวมถึงการระบุว่า คำบรรยาย คำอธิบายใดสมเหตุสมผลหรือไม่อย่างไร คำคาดการณ์จะเป็นไปได้หรือไม่ด้วยเหตุผลอะไร *องค์ประกอบที่ 2 การประเมินและออกแบบกระบวนการสืบเสาะหาความรู้ทางวิทยาศาสตร์ (Evaluate and Design Scientific Enquiry)* โดยบุคคลต้องสามารถประเมินข้อค้นพบทางวิทยาศาสตร์อย่างมีวิจารณญาณ แยกแยะคำถามทางวิทยาศาสตร์ว่า คำถามใดสามารถตอบได้ด้วยการสำรวจทางวิทยาศาสตร์ ใช้ความรู้เกี่ยวกับลักษณะสำคัญของการสำรวจตรวจสอบทางวิทยาศาสตร์ รวมถึงการเข้าใจถึงความสำคัญของการตั้งข้อสงสัยในการรายงานที่ปรากฏในสื่อหรือข้อค้นพบจากงานวิจัยต่าง ๆ ในแง่มุมที่อาจมีความคลุมเครือ การสรุปไม่สมเหตุสมผล ไม่มีข้อมูลมากพอหรือมีความลำเอียงได้ และ *องค์ประกอบที่ 3 การแปลความหมายข้อมูลและการใช้ประจักษ์พยานในเชิงวิทยาศาสตร์ (Interpret Data and Evidence Scientifically)* โดยบุคคลต้องแสดงออกถึงความสามารถในการตีความข้อมูลและหลักฐานทางวิทยาศาสตร์ที่ใช้ในการสร้างคำกล่าวอ้างหรือลงข้อสรุป นำเสนอข้อมูลที่ได้รับในรูปแบบอื่น ๆ เช่น ใช้คำพูดของตนเอง แผนภาพ หรือการแสดงอื่น ๆ แทนได้ ใช้เครื่องมือทางคณิตศาสตร์ในการวิเคราะห์หรือสรุปข้อมูลและใช้ความสามารถในการใช้วิธีการพื้นฐานของประจักษ์พยาน ข้อมูล หรือประเมินข้อสรุปที่ผู้อื่นสร้างขึ้นว่าสอดคล้องกับประจักษ์พยานที่มีหรือไม่ รวมถึงสามารถโต้แย้งอย่างสมเหตุสมผล

นิยามศัพท์เฉพาะ

1. *ระดับความมั่นใจ* หมายถึง ระดับที่ใช้คำถามวัดความมั่นใจในการตอบคำตอบของระดับเนื้อหา ประกอบด้วยมาตรประมาณค่าความมั่นใจ (Confidence Rating) ที่มีตัวเลือกเป็นระดับความมั่นใจที่มีแนวคิดมาจากมาตรประมาณค่า (Rating Scale) โดยในแบบสอบฉบับที่ 1 เป็นแบบสอบที่มีระดับความมั่นใจ 2 ระดับ ได้แก่ มั่นใจ (1) และ ไม่มั่นใจ (0) ส่วนแบบสอบฉบับที่ 2 เป็นแบบสอบที่มีระดับความมั่นใจ 4 ระดับ ได้แก่ มั่นใจมาก (3) มั่นใจ (2) ไม่มั่นใจ (1) และ เด่า (0)

2. *วิธีการตอบและตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 2 ระดับ* หมายถึง วิธีการที่ผู้สอบจะตอบลงในกระดาษคำตอบที่มี 2 ส่วนคือ คำตอบระดับเนื้อหา (Content Tier) และคำตอบระดับความมั่นใจ (Confidence Tier) ที่ประกอบด้วยความมั่นใจ 2 ระดับ ได้แก่ มั่นใจ และไม่มั่นใจ วิธีการตอบและตรวจให้คะแนนแสดงดังตัวอย่าง

ข้อสอบข้อที่ 0 คำตอบที่ถูกต้องคือ ข้อ ข ผู้สอบตอบข้อ ข และระบุความมั่นใจ “มั่นใจมาก”
จะได้คะแนน 2 คะแนน

ข้อ	ตัวเลือก				ระดับความมั่นใจ		คะแนน
	ก	ข	ค	ง	มั่นใจ	ไม่มั่นใจ	
0		×			×		2

ข้อสอบข้อที่ 0 คำตอบที่ถูกต้องคือ ข้อ ข ผู้สอบตอบข้อ ข และระบุความมั่นใจ “ไม่มั่นใจ”
จะได้คะแนน 1 คะแนน

ข้อ	ตัวเลือก				ระดับความมั่นใจ		คะแนน
	ก	ข	ค	ง	มั่นใจ	ไม่มั่นใจ	
0		×				×	1

ข้อสอบข้อที่ 0 คำตอบที่ถูกต้องคือ ข้อ ข ถ้าผู้สอบตอบผิดคือ ไม่ได้เลือกข้อ ข จะได้ 0
คะแนน ไม่ว่าจะระบุความมั่นใจในการตอบระดับใด

ข้อ	ตัวเลือก				ระดับความมั่นใจ		คะแนน
	ก	ข	ค	ง	มั่นใจ	ไม่มั่นใจ	
0	×				×		0

3. วิธีการตอบและตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 4 ระดับ
หมายถึง วิธีการที่ผู้สอบจะตอบลงในกระดาษคำตอบที่มี 2 ส่วนคือ คำตอบระดับเนื้อหา (Content
Tier) และคำตอบระดับความมั่นใจ (Confidence Tier) ที่ประกอบด้วยความมั่นใจ 4 ระดับ ได้แก่
มั่นใจมาก มั่นใจ ไม่มั่นใจ และเดา วิธีการตอบและตรวจให้คะแนนแสดงดังตัวอย่าง

ข้อสอบข้อที่ 0 คำตอบที่ถูกต้องคือ ข้อ ข ผู้สอบตอบข้อ ข และระบุความมั่นใจ “มั่นใจมาก”
จะได้คะแนน 4 คะแนน

ข้อ	ตัวเลือก				ระดับความมั่นใจ				คะแนน
	ก	ข	ค	ง	มั่นใจมาก	มั่นใจ	ไม่มั่นใจ	เดา	
0		×			×				4

ข้อสอบข้อที่ 0 คำตอบที่ถูกต้องคือ ข้อ ข ผู้สอบตอบข้อ ข และระบุความมั่นใจ “มั่นใจ” จะ
ได้คะแนน 3 คะแนน

ข้อ	ตัวเลือก				ระดับความมั่นใจ				คะแนน
	ก	ข	ค	ง	มั่นใจมาก	มั่นใจ	ไม่มั่นใจ	เดา	
0		×				×			3

ข้อสอบข้อที่ 0 คำตอบที่ถูกต้องคือ ข้อ ข ผู้สอบตอบข้อ ข และระบุความมั่นใจ “ไม่มั่นใจ”
จะได้คะแนน 2 คะแนน

ข้อ	ตัวเลือก				ระดับความมั่นใจ				คะแนน
	ก	ข	ค	ง	มั่นใจมาก	มั่นใจ	ไม่มั่นใจ	เดา	
0		×					×		2

ข้อสอบข้อที่ 0 คำตอบที่ถูกต้องคือ ข้อ ข ผู้สอบตอบข้อ ข และระบุความมั่นใจ “เดา” จะได้
คะแนน 1 คะแนน

ข้อ	ตัวเลือก				ระดับความมั่นใจ				คะแนน
	ก	ข	ค	ง	มั่นใจมาก	มั่นใจ	ไม่มั่นใจ	เดา	
0		×						×	1

ข้อสอบข้อที่ 0 คำตอบที่ถูกต้องคือ ข้อ ข ถ้าผู้สอบตอบผิดคือ ไม่ได้เลือกข้อ ข จะได้ 0
คะแนน ไม่ว่าจะระบุความมั่นใจในการตอบระดับใด

ข้อ	ตัวเลือก				ระดับความมั่นใจ				คะแนน
	ก	ข	ค	ง	มั่นใจมาก	มั่นใจ	ไม่มั่นใจ	เดา	
0	×				×				0

4. คุณภาพของวิธีการตรวจให้คะแนนความรู้บางส่วน หมายถึง ผลที่ได้จากการวิเคราะห์
ประกอบด้วย 4 ด้าน ได้แก่ ความยากและอำนาจจำแนกของแบบสอบ ความตรงเชิงโครงสร้างของ
แบบสอบ ความเที่ยงของแบบสอบ และฟังก์ชันสารสนเทศของแบบสอบ ที่ได้จากการตรวจให้
คะแนนความรู้บางส่วน ที่มีระดับความมั่นใจที่แตกต่างกันคือ ความมั่นใจ 2 ระดับ และความมั่นใจ 4
ระดับ

5. *ความยากของแบบสอบ* หมายถึง สัดส่วนของจำนวนคนที่ตอบข้อสอบข้อนั้นถูก ซึ่งได้จากการวิเคราะห์แบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่ต่างกันคือ ความมั่นใจ 2 ระดับ และความมั่นใจ 4 ระดับ ด้วยโปรแกรมสำเร็จรูป IRTPRO4 Student และวิเคราะห์ด้วย Generalized Partial Credit Model (G-PCM)

6. *อำนาจจำแนกของแบบสอบ* หมายถึง ความสามารถของข้อสอบในการจำแนกความแตกต่างระหว่างกลุ่มของผู้สอบที่มีผลสัมฤทธิ์ต่างกัน ซึ่งได้จากการวิเคราะห์แบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่ต่างกันคือ ความมั่นใจ 2 ระดับ และความมั่นใจ 4 ระดับ ด้วยโปรแกรมสำเร็จรูป IRTPRO4 Student และวิเคราะห์ด้วย Generalized Partial Credit Model (G-PCM)

7. *ความตรงเชิงโครงสร้าง* หมายถึง ความสอดคล้องระหว่างแบบสอบกับโครงสร้างภายในของคุณลักษณะที่ต้องการวัดคือ สมรรถนะทางวิทยาศาสตร์ ซึ่งได้จากการวิเคราะห์องค์ประกอบเชิงยืนยัน (Confirmatory Factor Analysis) โดยการประมาณค่าพารามิเตอร์ของโมเดลใช้โปรแกรมสำเร็จรูป Mplus และเป็น การประมาณค่าความน่าจะเป็นสูงสุด (Maximum Likelihood) การตรวจสอบความสอดคล้องระหว่างโมเดลและข้อมูล โดยการพิจารณาค่าดัชนีไคสแควร์ (χ^2), ไคสแควร์ต่อองศาอิสระ (χ^2/df), RMSEA, CFI และ SRMR

8. *ความเที่ยงของแบบสอบ* หมายถึง ความสัมพันธ์ระหว่างคะแนนที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่ต่างกันคือ ความมั่นใจ 2 ระดับ และความมั่นใจ 4 ระดับ โดยใช้ความเที่ยงแบบความสอดคล้องภายใน วิธีสัมประสิทธิ์แอลฟาครอนบาค (Cronbach's Alpha) ด้วยโปรแกรมสำเร็จรูป SPSS

9. *ฟังก์ชันสารสนเทศของแบบสอบ* หมายถึง การอธิบายความสัมพันธ์ระหว่างความสามารถของผู้สอบและโอกาสในการเลือกตอบแต่ละรายการที่กำหนด วิเคราะห์ด้วย Generalized Partial Credit Model (G-PCM) โดยใช้โปรแกรมสำเร็จรูป IRTPRO4 Student

ประโยชน์ที่ได้รับจากการวิจัย

1. ทำให้ทราบวิธีการตรวจให้คะแนนความรู้บางส่วนที่มีคุณภาพ รวมถึงขยายองค์ความรู้เกี่ยวกับวิธีการตรวจให้คะแนนความรู้บางส่วน เพื่อแก้ปัญหาการเดาข้อสอบ และการให้สารสนเทศในการจำแนกผู้เรียน

2. เพื่อให้สารสนเทศแก่ผู้สอนในการพิจารณารูปแบบการสอบและวิธีการตรวจให้คะแนนที่เหมาะสมกับผู้เรียน

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

งานวิจัยเรื่อง การเปรียบเทียบคุณภาพของแบบสอบเลือกตอบสองระดับด้านสมรรถนะทางวิทยาศาสตร์ที่มีระดับความมั่นใจที่แตกต่างกัน โดยประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบแบบตรวจให้คะแนนความรู้บางส่วนแบบทั่วไป ผู้วิจัยได้ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง แบ่งเป็น 5 ตอน ได้แก่

- ตอนที่ 1 แนวคิดเกี่ยวกับแบบสอบเลือกตอบ
- ตอนที่ 2 แนวคิดเกี่ยวกับทฤษฎีการตอบสนองข้อสอบ
- ตอนที่ 3 แนวคิดเกี่ยวกับวิธีการตรวจให้คะแนนความรู้บางส่วน
- ตอนที่ 4 แนวคิดเกี่ยวกับสมรรถนะทางวิทยาศาสตร์
- ตอนที่ 5 กรอบแนวคิดการวิจัย

ตอนที่ 1 แนวคิดเกี่ยวกับแบบสอบเลือกตอบ

แบบสอบเลือกตอบ เป็นรูปแบบของข้อสอบที่นิยมใช้ในการวัดและประเมินผล เนื่องจากมีความเป็นปรนัยในการตรวจให้คะแนน ใช้เวลาในการตรวจน้อย สามารถนำผลการตอบไปวิเคราะห์ข้อมูลทางสถิติได้ และเหมาะสำหรับการสอบที่มีผู้สอบจำนวนมาก (โชติกา ภาชีผล, ณีฎฐภรณ์ หลาวทอง และกมลวรรณ ตังธนากานนท์, 2558) ในการนำเสนอแนวคิดเกี่ยวกับแบบสอบเลือกตอบ แบ่งออกเป็น 5 หัวข้อ ดังนี้ 1) รูปแบบของแบบสอบเลือกตอบ 2) ข้อดีและข้อจำกัดของการทดสอบด้วยแบบสอบแบบเลือกตอบ 3) การเดาคำตอบในแบบสอบเลือกตอบ 4) ปัจจัยที่ส่งผลต่อการเดาในแบบสอบเลือกตอบ 5) แนวทางการป้องกันการเดาคำตอบ และ 6) การทดสอบด้วยแบบสอบเลือกตอบระดับต่าง ๆ

1.1 รูปแบบของแบบสอบเลือกตอบ

แบบสอบเลือกตอบ เป็นแบบสอบที่ข้อความคำถามมีการกำหนดรายการตัวเลือกให้กับผู้ตอบ โดยผู้ตอบสามารถใช้เวลาส่วนใหญ่ในการอ่าน คิด และเลือกรายการคำตอบจากรายการตัวเลือก แบบสอบเลือกตอบ แบ่งเป็น 3 รูปแบบ ได้แก่ 1) แบบสอบเลือกตอบแบบถูก-ผิด (True-False) เป็นข้อสอบที่คำตอบมีความเป็นไปได้สองอย่าง เช่น ถูก-ผิด ใช่-ไม่ใช่ 2) แบบสอบเลือกตอบแบบจับคู่ (Matching) เป็นข้อสอบที่ให้ผู้สอบจับคู่คำตอบ และ 3) แบบสอบเลือกตอบแบบหลายตัวเลือก

(Multiple-Choice) เป็นข้อสอบที่ให้ผู้สอบเลือกคำตอบจากคำตอบที่โจทย์กำหนด (ศิริชัย กาญจนวาสี, 2555)

รูปแบบของแบบสอบที่นิยมนำไปใช้ในการทดสอบเพื่อวัดความรู้ของผู้สอบมากที่สุดคือ แบบสอบเลือกตอบแบบหลายตัวเลือก เพราะสามารถวัดความรู้ได้ตั้งแต่ระดับความจำ ความเข้าใจ ไปจนถึงระดับที่สูงขึ้น อีกทั้งยังสามารถใช้เครื่องจักรในการตรวจให้คะแนนได้อย่างรวดเร็วและแม่นยำ มีค่าใช้จ่ายต่ำ สามารถควบคุมเวลาในการบริหารจัดการได้ และสามารถวัดและประเมินผลได้ครอบคลุมตรงตามเนื้อหามากกว่าแบบสอบรูปแบบอื่น (ศิริชัย กาญจนวาสี, 2555; อรพรรณ แก้วน้อย, 2557)

ข้อจำกัดของแบบสอบเลือกตอบคือ ในการสร้างข้อสอบแบบเลือกตอบให้มีคุณภาพนั้นเป็นไปได้ยาก ต้องอาศัยผู้ทรงคุณวุฒิที่มีความเชี่ยวชาญทางด้านเนื้อหาและมีทักษะในการสร้างแบบสอบเป็นอย่างดี นอกจากนี้ยังมีข้อจำกัดในการนำไปใช้ในเชิงปฏิบัติคือ 1) ไม่สามารถหลีกเลี่ยงการเดาคำตอบของผู้สอบที่มีความรู้ไม่เพียงพอ ทำให้ความแปรปรวนของคะแนนและความคลาดเคลื่อนที่วัดได้จากแบบสอบจึงอาจมีแนวโน้มสูง ส่งผลให้ความเที่ยงและความตรงต่ำ (Budesu & Bo, 2015) และ 2) ไม่สามารถวัดผู้ตอบที่มีความรู้บางส่วน เนื่องจากการตรวจให้คะแนนมีเพียง 2 ค่าคือ 0 (ผิด/ไม่มีความรู้) กับ 1 (ถูก/มีความรู้) จึงไม่เพียงพอต่อการจำแนกผู้สอบกลุ่มนี้ออกจากกลุ่มที่มีความรู้สมบูรณ์ และกลุ่มที่ไม่มีความรู้เลย (Bo et al., 2015)

จากที่กล่าวมาข้างต้น การเดา จึงเป็นสิ่งที่ส่งผลกระทบต่อการวัดและประเมินผลมากที่สุด เนื่องจากการเปิดช่องให้ผู้สอบที่มีความรู้ไม่เพียงพอมีโอกาสได้คะแนนจากการตอบคำถาม คะแนนที่วัดได้ไม่สะท้อนถึงความรู้จริงของผู้สอบ จึงอาจเกิดข้อผิดพลาดในการนำผลคะแนนไปใช้เพื่อประเมินหรือตัดสินใจ โดยเฉพาะเรื่องที่มีความสำคัญและไม่มีโอกาสติดตามตรวจสอบได้อีก เช่น การสอบเพื่อสำเร็จการศึกษา การสอบคัดเลือกเพื่อเข้าศึกษาต่อหรือทำงาน (ศิริชัย กาญจนวาสี, 2555)

1.2 ข้อดีและข้อจำกัดของการทดสอบด้วยแบบสอบแบบเลือกตอบ

ข้อดีของการทดสอบด้วยแบบสอบแบบเลือกตอบ มีดังนี้

- 1) สามารถทดสอบเนื้อหาหลาย ๆ ประเภทได้ในเวลาที่รวดเร็ว
- 2) สามารถใช้วัดทักษะทางพุทธิพิสัยและทักษะการเรียนรู้ในหลาย ๆ ระดับได้
- 3) การตรวจสอบให้คะแนนมีความเป็นปรนัยและมีความเที่ยงสูง
- 4) ตรวจให้คะแนนได้ง่ายและรวดเร็ว
- 5) เป็นวิธีการที่เหมาะสมสำหรับนักเรียนที่มีทักษะการเขียนต่ำ
- 6) มีความเหมาะสมในการวิเคราะห์คุณภาพข้อสอบเป็นรายข้อเมื่อมีการกำหนด

คุณลักษณะในการวัดหลายคุณลักษณะ

7) ได้ข้อมูลผลการวิจัยที่มีคุณค่าและทันท่วงทีที่สามารถนำไปใช้สัมภาษณ์เพื่อค้นหาข้อมูลวิจัยเชิงลึก

ข้อจำกัดของการทดสอบด้วยแบบสอบแบบเลือกตอบ มีดังนี้

- 1) คำตอบที่นักเรียนตอบถูกอาจเกิดจากการเดา
- 2) แบบสอบแบบเลือกตอบเพียงอย่างเดียวไม่สามารถอธิบายความเข้าใจในเชิงลึกของนักเรียนได้
- 3) นักเรียนถูกบังคับให้เลือกตอบในตัวเลือกที่มี ซึ่งนักเรียนอาจไม่ได้มีความคิดตรงกับตัวเลือกที่มีให้
- 4) การสร้างแบบสอบแบบเลือกตอบให้มีคุณภาพทำได้ยาก

1.3 การเดาคำตอบในแบบสอบเลือกตอบ

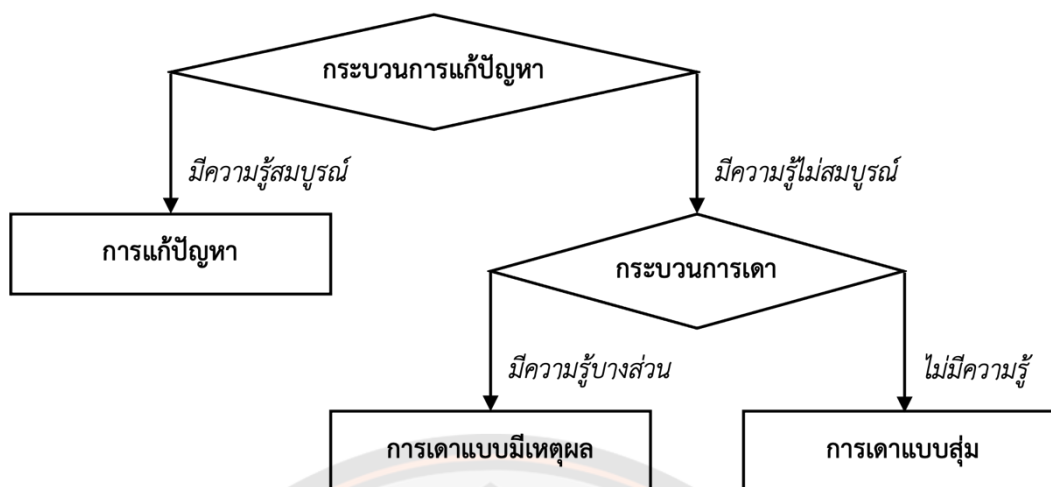
จากการศึกษาวิจัยที่เกี่ยวข้องกับการเดาคำตอบ พบว่ามีการให้ความหมายและรูปแบบไว้ดังนี้

Birnbaum (1968) ให้ความหมายในเชิงของการวัดผลว่า การเดา หมายถึง โอกาสในการตอบข้อสอบถูกของผู้ตอบที่ไม่มีความรู้ใดเลยเกี่ยวกับข้อคำถาม

Han (2012) ให้ความหมาย การเดา ว่าหมายถึง สิ่งที่เกิดขึ้นเมื่อผู้ตอบไม่ทราบคำตอบที่ถูกต้องของข้อคำถาม แต่ยังพยายามที่จะตอบคำถามนั้นเพื่อต้องการเพิ่มโอกาสในการได้คะแนน

Mehrens และ Lehmann (1984) แบ่งการเดาเป็น 2 รูปแบบ ได้แก่ 1) การเดาแบบสุ่ม (Blind-, Wild-, Lucky-, Random Guessing) เป็นการสุ่มเลือกคำตอบจากบรรดาตัวเลือกทั้งหมดที่เกิดขึ้นเมื่อผู้ตอบไม่มีความรู้ใดเลยเกี่ยวกับคำถาม และ 2) การเดาแบบมีเหตุผล (Informed-, Educated-, Logical-, Ability- Based Guessing) เป็นการเลือกคำตอบโดยอาศัยการตัดตัวเลือกที่คาดว่าจะผิดออกไปก่อนจากนั้นจึงเดาแบบสุ่มจากตัวเลือกที่เหลืออยู่เกิดขึ้นเมื่อผู้ตอบมีความรู้บางส่วนเกี่ยวกับคำถาม

Hutchinson (1991) อธิบายว่า ในการตอบข้อคำถามจะประกอบด้วย 2 กระบวนการ ได้แก่ 1) กระบวนการแก้ปัญหา เป็นการตั้งใจพยายามหาคำตอบ และ 2) กระบวนการเดา เป็นการคาดเดาคำตอบเมื่อผู้ตอบมีความรู้ไม่สมบูรณ์ แบ่งเป็น 2 รูปแบบย่อย คือ 2.1) การเดาแบบมีเหตุผล เกิดขึ้นเมื่อผู้ตอบมีความรู้บางส่วนเกี่ยวกับข้อคำถาม และ 2.2) การเดาแบบสุ่ม เกิดขึ้นเมื่อผู้ตอบไม่มีความรู้ใดเลยเกี่ยวกับข้อคำถาม ดังภาพ 1



ภาพ 1 โมเดลอธิบายการเดาตามแนวคิดของ

ที่มา: Hutchinson (1991)

รูปแบบการเดาทั้งสองแบบนี้มีความแตกต่างกันในการให้สารสนเทศเกี่ยวกับความรู้ของผู้ตอบ กล่าวคือ การเดาแบบสุ่มจะถือว่าไม่มีการให้สารสนเทศใดทั้งสิ้น เนื่องจากการเดาไม่ได้มาจากการใช้ความรู้ของผู้ตอบ แตกต่างจากการเดาแบบมีเหตุผลที่ถือว่ามีสารสนเทศบางส่วน เนื่องจากการเดามีการนำความรู้บางส่วนของผู้ตอบเข้ามาช่วยในการเดาด้วย (Smith, 1993)

ดังนั้นจึงสามารถสรุปได้ว่า การเดา เป็นพฤติกรรมในการตอบข้อความของผู้ตอบ เมื่อผู้ตอบมีความรู้ไม่สมบูรณ์ โดยแบ่งได้ 2 รูปแบบ คือ 1) การเดาแบบสุ่ม เกิดขึ้นเมื่อผู้ตอบไม่มีความรู้ใดเลยเกี่ยวกับข้อความ จึงถือว่าไม่มีการให้สารสนเทศใดเกี่ยวกับความรู้ของผู้ตอบ และ 2) การเดาแบบมีเหตุผล เกิดขึ้นเมื่อผู้ตอบมีความรู้บางส่วนเกี่ยวกับข้อความ จึงมีการให้สารสนเทศบางส่วนเกี่ยวกับความรู้ของผู้ตอบ

1.4 ปัจจัยที่ส่งผลต่อการเดาในแบบสอบถามเลือกตอบ

จากการศึกษางานวิจัยที่เกี่ยวข้องสามารถสังเคราะห์ปัจจัยที่ส่งผลต่อการเดา (อิริตม์ สุขสกุลวัฒน์, 2563) แบ่งได้เป็น 2 ด้าน ดังนี้

1) คุณลักษณะของข้อสอบ

1.1) ความยาก เมื่อข้อสอบมีระดับความยากสูงกว่าความรู้ของผู้ตอบ ผู้ตอบจะเริ่มมีแนวโน้มคาดเดาคำตอบมากขึ้น โดยข้อสอบระดับง่ายมักเปิดช่องให้มีโอกาสในการเดา ถูกสูงกว่าปกติ ตรงกันข้ามกับข้อสอบระดับยาก ซึ่งพบว่าโอกาสในการเดาถูกมักมีแนวโน้มลดต่ำลง

(Tsai & Lin, 2015) นอกจากนี้ตำแหน่งของข้อสอบที่มีระดับความยากแตกต่างกันก็อาจส่งผลต่อการเดาเช่นกัน ตัวอย่างเช่น หากข้อสอบในช่วงเริ่มต้นยากเกินไป ผู้ตอบอาจหมดกำลังใจในการสอบและเริ่มเดาคำตอบตั้งแต่วางแรก หรือถ้าข้อสอบในช่วงท้ายยากมากขึ้นและผู้ตอบรู้สึกเหนื่อยล้า ผู้ตอบก็อาจเริ่มเดาคำตอบเช่นกัน (Cao & Stokes, 2008)

1.2) *อำนาจจำแนก* ข้อสอบที่มีอำนาจจำแนกต่ำหรือไม่สามารถจำแนกผู้ตอบที่มีความสามารถสูงและต่ำออกจากกันได้มักเปิดช่องให้ผู้ตอบทั้งสองกลุ่มมีโอกาสในการตอบถูกใกล้เคียงกัน และการที่ผู้ตอบกลุ่มต่ำมีโอกาสในการตอบถูกใกล้เคียงกับผู้ตอบกลุ่มสูงย่อมหมายความว่า ข้อสอบข้อนั้นมีโอกาสในการเดาถูกสูงหรือเดาถูกได้ง่าย (Hambleton et al., 1991)

1.3) *จำนวนตัวเลือก* หากจำนวนตัวเลือกในข้อสอบนั้นมีมากขึ้น โอกาสในการเดาถูกย่อมต่ำลง สังเกตได้จากโอกาสในการเดาถูกแบบสุ่ม ซึ่งมีค่า $1/m$ เมื่อ m คือจำนวนตัวเลือกของข้อสอบ (Han, 2012)

1.4) *คุณภาพตัวลวง* ตัวลวงที่มีคุณภาพจะมีผลต่อการเดาเป็นอย่างมาก โดยเฉพาะในกลุ่มผู้ตอบความรู้ต่ำ ซึ่งพบว่า มีแนวโน้มสามารถล่อลวงผู้ตอบความรู้ต่ำหรือมีความรู้ไม่สมบูรณ์ได้ดี การเดาของผู้ตอบกลุ่มนี้จึงมักไม่ใช่การเดาแบบสุ่ม (Lord, 1974)

1.5) *วิชาที่สอบ/ลักษณะตัวเลือก* โดยทั่วไปในวิชาคณิตศาสตร์ ตัวเลือกในข้อสอบมักมีลักษณะเป็นตัวเลข ซึ่งมีจุดเด่นคือ ผู้ตอบที่มีความรู้สมบูรณ์มักตอบข้อสอบได้ทั้งหมด แต่หากไม่มีความรู้หรือมีความรู้เพียงบางส่วนจะตอบไม่ได้เลย (All-or-None) การเดาในวิชาคณิตศาสตร์จึงมักเป็นการเดาแบบสุ่ม ส่วนในวิชาทางภาษา ตัวเลือกในข้อสอบมักจะมีลักษณะเป็นข้อความ ผู้ตอบที่มีความรู้ไม่เพียงพอในการตอบสามารถอ่านข้อความในแต่ละตัวเลือกและใช้ความรู้ที่มีตัดตัวเลือกที่คาดว่าจะไม่ถูกออกไปก่อน แล้วจึงเดาแบบสุ่มจากตัวเลือกที่เหลือ การเดาในวิชาทางภาษาจึงมักเป็นการเดาแบบมีเหตุผล (San Martin et al., 2006)

1.6) *วิธีการตอบ/การตรวจให้คะแนน* ข้อสอบแบบเลือกตอบที่มีตัวถูกหลายตัว (Multiple Correct Options) และมีการตรวจให้คะแนนบางส่วน (Partial Credit) พบว่าสามารถช่วยลดโอกาสในการเดาถูกของผู้ตอบลงไปได้มาก เมื่อเทียบกับข้อสอบเลือกตอบที่มีตัวถูกเพียงตัวเดียว (Ben-Simon et al., 1997)

2) คุณลักษณะของผู้สอบ

2.1) *เพศ* การศึกษาความแตกต่างของพฤติกรรมสอบระหว่างเพศชายกับเพศหญิงในบริบทการสอบที่มีความสำคัญสูง (High-stakes Testing) เช่น การสอบเลื่อนชั้น การสอบคัดเลือกเพื่อเข้าศึกษาต่อ มักพบว่า เพศชายมีพฤติกรรมชอบเสี่ยงเดาคำตอบมากกว่าเว้นคำตอบไว้ ตรงกันข้ามกับเพศหญิงที่ไม่ชอบเสี่ยงเดาคำตอบแม้จะไม่มีทางเลือกเมื่อเดาผิดก็ตาม (Baldiga, 2014) ทั้งนี้ ผลดังกล่าวอาจมาจากความแตกต่างระหว่างเพศเกี่ยวกับแนวคิดความเป็นเลิศ

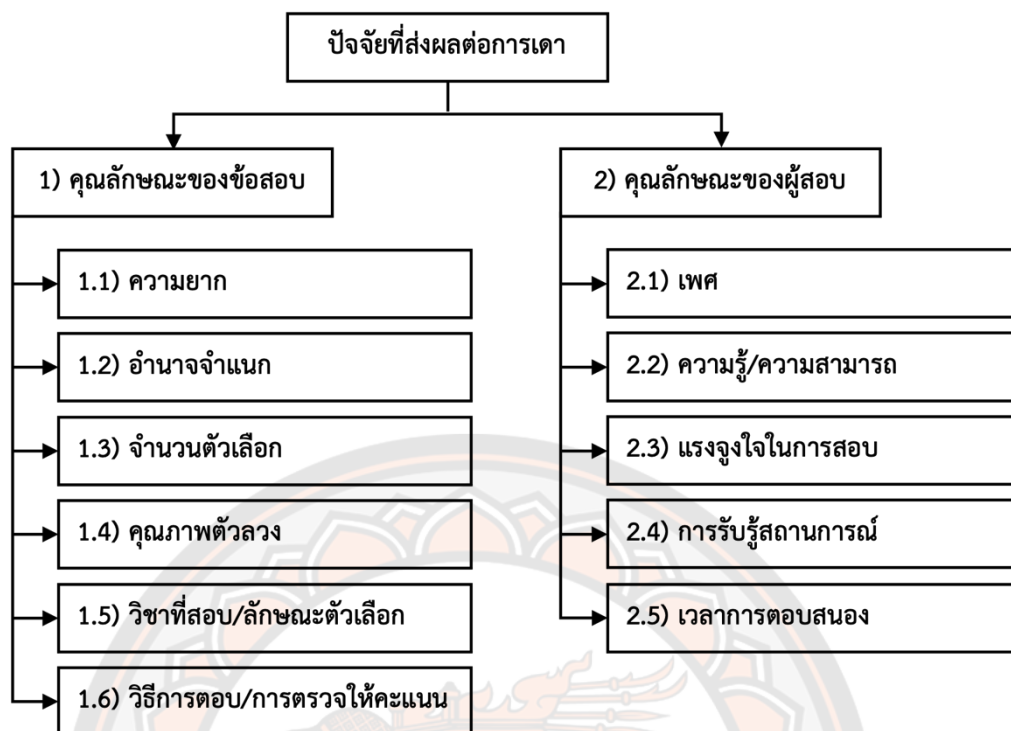
ในการแข่งขันในสถานการณ์ที่มีความเสี่ยง โดยเพศชายอาจคิดว่าการได้ผลงานดีคือการเพิ่มคะแนนสอบให้ได้สูงสุดหรือการไม่ยอมรับว่าไม่ทราบคำตอบจึงตัดสินใจเสี่ยงเดาคำตอบมากกว่าเว้นคำตอบ ส่วนเพศหญิงอาจคิดว่าการได้ผลงานดีคือการไม่ต้องรับการลงโทษใดจึงตัดสินใจเว้นคำตอบมากกว่าเสี่ยงเดาคำตอบ (Croson & Gneezy, 2009)

2.2) *ความรู้/ความสามารถ* ผู้ตอบความสามารถต่ำมักถูกล่อลวงด้วยตัวลวงที่มีคุณภาพ ทำให้โอกาสในการเดาถูกมีแนวโน้มต่ำกว่าโอกาสในการเดาถูกแบบสุ่ม (Lord, 1974) ส่วนผู้ตอบความสามารถสูงมักใช้ความรู้ที่มีตัดตัวเลือกที่คาดว่าจะไม่ถูกต้องออกไปก่อน แล้วจึงเดาคำตอบจากตัวเลือกที่เหลือ โอกาสในการเดาถูกจึงมีแนวโน้มสูงกว่าโอกาสในการเดาถูกแบบสุ่ม (San Martin et al., 2006)

2.3) *แรงจูงใจในการสอบ* แรงจูงใจในการสอบมักมีความสัมพันธ์กับความสำคัญของการสอบ และอาจส่งผลต่อการเดา โดยเฉพาะการสอบที่มีความสำคัญต่ำ (Low-stakes Testing) เช่น การสอบเพื่อติดตามพัฒนาของผู้เรียน การสอบประเมินระดับชาติที่ไม่มีผลได้ผลเสียต่อตัวผู้สอบโดยตรง เป็นต้น การสอบในลักษณะนี้ผู้ตอบมักให้ความสนใจและพยายามทำข้อสอบในช่วงแรกเท่านั้น เนื่องด้วยความอยากรู้อยากเห็น แต่พอถึงจุดหนึ่งผู้ตอบจะละทิ้งความพยายามและเริ่มเดาคำตอบ นอกจากนี้การสอบดังกล่าวมักใช้ข้อสอบที่มีจำนวนมากและมีรูปแบบการตอบที่หลากหลาย ผู้ตอบจึงเกิดความเมื่อยล้าจากการทำข้อสอบได้ง่าย เป็นผลทำให้เกิดการเดามากขึ้นในช่วงท้ายของการทดสอบ (Cao & Stokes, 2008)

2.4) *การรับรู้สถานการณ์* เมื่อมีการลงโทษหรือหักคะแนนจากการเดาผิด ผู้ตอบที่คาดการณ์ได้ว่ามีคะแนนสูงกว่าคะแนนที่ยอมรับได้ต่ำสุด (Minimal Acceptable Point) หรือคะแนนอ้างอิง มักจะเว้นคำตอบในข้อที่ไม่แน่ใจไว้เพราะเชื่อมั่นว่ายังได้คะแนนอยู่ในตำแหน่งเดิม จึงไม่จำเป็นต้องเสี่ยงเดาคำตอบ ส่วนผู้ตอบที่คาดการณ์ว่ามีคะแนนต่ำกว่าคะแนนอ้างอิง มักมีแนวโน้มเดาคำตอบมากกว่า เนื่องจากต้องเสี่ยงเพื่อเพิ่มโอกาสให้ได้อยู่ในตำแหน่งที่สูงกว่าคะแนนอ้างอิง (Bareby-Meyer, Meyer, & Flascher, 2002)

2.5) *เวลาการตอบสนอง* เมื่อใช้การสอบผ่านทางคอมพิวเตอร์ (Computer-based Test) เวลาในการตอบสนองข้อสอบอาจเข้ามามีบทบาทเป็นตัวจำแนกความแตกต่างระหว่างกระบวนการแก้ปัญหา กับกระบวนการเดาด้วยการพิจารณาจากขีดแบ่งเวลา (Time Threshold) ที่ใช้ในการทำข้อสอบแต่ละข้อ โดยการตอบที่ใช้เวลาน้อยกว่าขีดแบ่งเวลามีแนวโน้มสูงที่จะเป็นการเดาคำตอบ ส่วนการตอบที่ใช้เวลามากกว่าขีดแบ่งเวลาจะมีแนวโน้มเป็นการแก้ปัญหา มากกว่าการเดา (Wise & Kong, 2005) รายละเอียดโดยสรุปดังภาพ 2



ภาพ 2 ปัจจัยที่ส่งผลต่อการเดาในแบบสอบเลือกตอบ

1.5 แนวทางการป้องกันการเดาคำตอบ

แนวทางการป้องกันการเดาคำตอบ แบ่งได้เป็น 4 วิธี (ธีรุตม์ สุขสกุลวัฒน์, 2563) ได้แก่ 1) การปรับวิธีการตรวจให้คะแนน 2) การเปลี่ยนรูปแบบการตอบ 3) การเปลี่ยนโครงสร้างของข้อสอบ และ 4) การหาความเชื่อมั่นของผลการตอบ มีรายละเอียดดังนี้

1) การปรับวิธีการตรวจให้คะแนน

1.1) การตรวจให้คะแนนตามสูตร (Formula Scoring) เช่น การใช้สูตรปรับแก้การเดา (Correction for Guessing) หรือ $C = R - W/(m - 1)$ เมื่อ C คือ คะแนนปรับแก้ R คือ จำนวนข้อที่ตอบถูก W คือ จำนวนข้อที่ตอบผิด และ m คือ จำนวนตัวเลือกในข้อสอบ วิธีนี้พบว่าสามารถลดแรงจูงใจในการเดาของผู้สอบกลุ่มที่ไม่ชอบความเสี่ยงได้ด้วยการเว้นคำตอบแทนการเดา (Holzinger, 1924) ซึ่งอ่าวายสนับสนุนคุณธรรมและจริยธรรมของผู้สอบ (Lindquist & Hoover, 2015) และอาจช่วยเพิ่มความเที่ยงให้สูงขึ้น (Budescu & Bar-Hillel, 1993) อย่างไรก็ตาม วิธีนี้มีข้อจำกัดคือ การลงโทษหรือการหักคะแนนเมื่อเดาผิดจะส่งผลเสียทั้งต่อตัวผู้สอบและผู้ให้ผลคะแนน เนื่องจากในสถานการณ์จริงอาจมีผู้สอบหลายประเภท เช่น กลุ่มที่ไม่ชอบเสี่ยง กลุ่มที่ไม่ชอบสูญเสีย และกลุ่มอนุรักษ์นิยมการทำข้อสอบแบบดั้งเดิม ผู้สอบแต่ละกลุ่มอาจถูกลงโทษอย่างไม่เป็นสัดส่วน ส่งผลให้คะแนนที่ประมาณได้เกิดความลำเอียง (Budescu & Bo, 2015)

1.2) *การกำหนดน้ำหนักให้ข้อสอบ (Item Weighting)* เช่น การกำหนดให้ข้อสอบแต่ละข้อมีน้ำหนักคะแนนแตกต่างกัน ซึ่งอาจกำหนดตามระดับความยาก ความตรง ความสามารถในการวินิจฉัย ด้วยเทคนิคการวิเคราะห์ถดถอยหรือการวิเคราะห์องค์ประกอบ (Sykes & Hou, 2003) อย่างไรก็ตามวิธีนี้ยังคงค่อนข้างมีความซับซ้อนและกำหนดน้ำหนักที่เหมาะสมได้ยาก

1.3) *การกำหนดน้ำหนักให้ตัวเลือก (Option Weighting)* เช่น การกำหนดให้ตัวเลือกมีความรุนแรงในการลงโทษแตกต่างกันเมื่อผู้สอบเดาผิด ซึ่งสามารถกำหนดได้หลายวิธี เช่น กำหนดตามการตัดสินใจของผู้เชี่ยวชาญจากโครงสร้างความรู้ที่มุ่งวัด (Smith, 1987) กำหนดจากผลการตอบที่ได้จากแบบสอบคู่ขนานที่เป็นมาตรฐานหรือความรู้ก่อนหน้าเกี่ยวกับน้ำหนักคะแนน (Ben-Simon et al., 1997) วิธีดังกล่าวพบว่าอาจช่วยเพิ่มความเที่ยงแบบสอดคล้องภายใน (Internal Consistency Reliability) ให้มีค่าสูงขึ้นได้ (Echternacht, 1976) อย่างไรก็ตามวิธีนี้ไม่ได้ได้รับความนิยม เนื่องจากมีค่าใช้จ่ายสูงในการวางแผนกำหนดน้ำหนัก และมักพบปัญหาการกำหนดน้ำหนักให้มีความเหมาะสม รวมถึงการคำนวณคะแนนยังมีความซับซ้อนและไม่สะดวกต่อการอธิบาย ขั้นตอนการตรวจให้ผู้สอบทราบ (Frary, 1989)

2) การเปลี่ยนรูปแบบการตอบ

2.1) *การตอบปฏิเสธตัวเลือกที่ผิด* เช่น การสอบแบบตัดทิ้ง (Elimination Testing) การตอบรูปแบบนี้ผู้สอบต้องระบุว่าตัวเลือกใดเป็นตัวเลือกที่ผิดบ้าง ส่วนตัวเลือกที่คิดว่าเป็นตัวเลือกที่ถูกหรือไม่แน่ใจให้เว้นไว้ กล่าวคือ ผู้สอบต้องตัดตัวเลือกที่คิดว่าผิดออกไปให้ได้มากที่สุดเท่าที่สามารถทำได้ การตรวจให้คะแนนจะพิจารณาจากทุกตัวเลือกที่ทำเครื่องหมายตอบ และคำนวณจากจำนวนตัวเลือกที่ตัดได้ถูกและผิด (Bush, 2015)

2.2) *การตอบตัวเลือกที่มีแนวโน้มถูก* เช่น การสอบแบบเลือกบางส่วน (Subset Selection Testing) การตอบรูปแบบนี้ผู้สอบสามารถเลือกตัวเลือกที่คาดว่าเป็นตัวเลือกที่ถูกต้องได้หลายตัว แต่ตัวเลือกที่ถูกจะมีเพียงตัวเลือกเดียวเท่านั้น การตรวจให้คะแนนจะมีการลดหลั่นตามจำนวนตัวเลือกที่ผู้สอบเลือกตอบมากขึ้น เนื่องจากเชื่อว่า การตอบตัวเลือกมากขึ้นแสดงถึงความรู้ของผู้สอบที่มีค่าน้อยลง (Dressel & Schmidt, 1953)

รูปแบบการตอบทั้งสองมีความคล้ายคลึงกันในการวัดความรู้บางส่วนของผู้สอบ และการลงโทษเมื่อเดาผิดส่งผลต่อการลดโอกาสในการเดาถูก ช่วยจำแนกผู้สอบที่มีระดับความรู้แตกต่างกันได้ชัดเจน และยังเป็น การเพิ่มความเที่ยงได้ทางหนึ่ง (Gibbons, Olkin, & Sobel, 1977) อย่างไรก็ตาม การตอบรูปแบบนี้มีข้อจำกัดเกี่ยวกับการตอบและการตรวจให้คะแนน ซึ่งต้องอาศัยเวลานานกว่าการสอบทั่วไป และแม้ว่าการตอบทั้งสองจะมีความเท่าเทียมกันในเชิงคณิตศาสตร์ แต่ในเชิงกลวิธีการคิดเพื่อตอบคำถามยังคงมีความแตกต่างกันอยู่ (Bereby-Meyer, Meyer, & Budescu, 2003)

3) การเปลี่ยนโครงสร้างข้อสอบ

วิธีการเปลี่ยนโครงสร้างข้อสอบ เช่น การปรับข้อสอบให้มีตัวถูกหลายตัว ผู้สอบต้องเลือกตัวเลือกที่ถูกต้องให้ครบทุกตัว โดยไม่ทราบว่ามีตัวเลือกที่ถูกต้องมีจำนวนเท่าใด (Frisbie, 1992) ข้อดีคือ โอกาสในการเดาถูกจะลดลงอย่างมากเมื่อเทียบกับข้อสอบที่มีตัวถูกเพียงตัวเดียว และพบว่ามีแนวโน้มให้ความเที่ยงและความตรงสูงกว่าข้อสอบที่มีตัวถูกเพียงตัวเดียวด้วย (Hsu, Moss, & Choosak, 1984) อย่างไรก็ตาม ข้อสอบดังกล่าวมีข้อจำกัดเกี่ยวกับจำนวนตัวถูกในแต่ละข้อซึ่งอาจมีไม่เท่ากัน คะแนนในแต่ละข้อจึงไม่เท่ากันด้วย หากจะให้คะแนนแต่ละข้อเท่ากันก็ต้องใช้สูตรปรับแก้ จึงเป็นการเพิ่มภาระในการตรวจและไม่สะดวกต่อการบริหารจัดการ (Ben-Simon et al., 1997)

4) การสร้างความเชื่อมั่นให้ผลการตอบ

วิธีการสร้างความเชื่อมั่นให้ผลการตอบ เช่น การเปรียบเทียบระหว่างความรู้จากการประเมินตนเองกับความรู้ที่วัดได้จากแบบสอบ วิธีการนี้เมื่อผู้สอบเลือกคำตอบที่คิดว่าถูกแล้วต้องระบุด้วยว่ามั่นใจแค่ไหนว่าผลการตอบนั้นถูกต้อง เพื่อเป็นการสร้างความเชื่อมั่นว่าผลการตอบมีความสอดคล้องกับความรู้จริงของผู้สอบ ข้อดีของวิธีนี้พบว่าอาจให้ความเที่ยงสูงกว่าเมื่อเทียบกับการตรวจให้คะแนนแบบ 0 หรือ 1 และการตรวจให้คะแนนตามสูตร (Hambleton, Roberts, & Traub, 1970) อย่างไรก็ตาม วิธีนี้พบว่าอาจมีปัจจัยที่ไม่เกี่ยวกับความรู้ (Non-cognitive Factor) เข้ามาเกี่ยวข้องกับการหาความเชื่อมั่น การวัดจึงอาจมีความคลาดเคลื่อนสูง อีกทั้งยังต้องอาศัยเวลาในการบริหารจัดการที่มากขึ้นและมีวิธีการตรวจให้คะแนนที่ซับซ้อนมากกว่าปกติ (Jacobs, 1971) นอกจากนี้ยังควรระมัดระวังหากมีการให้ข้อมูลผิด (Misinformation) ตัวอย่างเช่น ผู้สอบมั่นใจว่าคำตอบถูก แต่ปรากฏว่าคำตอบนั้นผิด ซึ่งกรณีดังกล่าวอาจส่งผลทำให้ความเที่ยงต่ำลง (พัชรี มีวรรณ, 2540)

1.6 การทดสอบด้วยแบบสอบแบบเลือกตอบระดับต่าง ๆ

ธนบดี อินทาดกรวด (2560) ได้สรุปลักษณะของแบบสอบแบบเลือกตอบสองระดับ สามระดับ และสี่ระดับ ดังตาราง 1

ตาราง 1 ลักษณะของแบบสอบแบบเลือกตอบสองระดับ สามระดับ และสี่ระดับ

แบบสองระดับ	แบบสามระดับ	แบบสี่ระดับ
1. การตอบระดับคำตอบ	1. การตอบระดับคำตอบ	1. การตอบระดับคำตอบ
2. การตอบระดับเหตุผล	2. การตอบระดับเหตุผล	2. ความมั่นใจในการตอบระดับคำตอบ
	3. ความมั่นใจในการตอบ	3. การตอบระดับเหตุผล
		4. ความมั่นใจในการตอบระดับเหตุผล

Gurel et al. (2015) ได้สรุปข้อดีและข้อจำกัดของแบบสอบแบบเลือกตอบสองระดับ สามระดับ และสี่ระดับ ดังตาราง 2

ตาราง 2 ข้อดีและข้อจำกัดของแบบสอบแบบเลือกตอบสองระดับ สามระดับ และสี่ระดับ

	แบบสองระดับ	แบบสามระดับ	แบบสี่ระดับ
ข้อดี	<ol style="list-style-type: none"> 1. ประหยัดเวลา 2. ตรวจสอบให้คะแนนได้รวดเร็ว 3. ตรวจสอบให้คะแนนได้อย่างเป็นปรนัย 4. มีหลักฐานที่น่าเชื่อถือสูงเกี่ยวกับความตรง 5. ใช้ได้กับผู้เข้าสอบจำนวนมาก 6. มีโอกาสที่จะตัดสินสัดส่วนของนักเรียนที่ตอบผิดพลาดเชิงบวกและผิดพลาดเชิงลบ 	<ol style="list-style-type: none"> 1. มีข้อดีเหมือนกับแบบสอบแบบเลือกตอบสองระดับ 2. มีการจำแนกการตอบที่ผิดออกเป็นกลุ่มที่ตอบผิดพลาดเนื่องเพราะ มโนทัศน์ที่คลาดเคลื่อน หรือเพราะความผิดพลาดในการตอบ หรือเพราะไม่มีความรู้ 	<ol style="list-style-type: none"> 1. มีข้อดีเหมือนกับแบบสอบแบบเลือกตอบสามระดับ 2. เป็นการประเมินที่แยกกลุ่มที่มี มโนทัศน์ที่คลาดเคลื่อนออกจากกลุ่มที่ตอบเพราะความคลาดเคลื่อนของการตอบ และกลุ่มที่ตอบเพราะไม่มีความรู้ได้อย่างชัดเจน
ข้อจำกัด	<ol style="list-style-type: none"> 1. มีการประมาณค่าที่เกินความเป็นจริงเมื่อไม่สามารถระบุกลุ่มที่ตอบผิดพลาด 	<ol style="list-style-type: none"> 1. มีการประมาณค่าที่ต่ำกว่าความเป็นจริง เพราะไม่สามารถตัดสินได้แน่ชัดใน 	<ol style="list-style-type: none"> 1. ใช้เวลาทดสอบนาน

แบบสองระดับ	แบบสามระดับ	แบบสี่ระดับ
เพราะขาดความรู้ได้อย่างแน่นอน	กลุ่มนักเรียนที่ตอบผิด เพราะไม่มีความรู้ในกรณีที่มีความมั่นใจ เพราะไม่สามารถระบุแน่ชัดว่านักเรียนมั่นใจในระดับหนึ่งหรือระดับสอง หรือมั่นใจทั้งสองระดับ	
	2. มีการประมาณค่าคะแนนที่ได้มากกว่าความเป็นจริง	

จากตาราง 2 พบว่า แบบสอบแบบเลือกตอบสี่ระดับ สามารถวินิจฉัยผู้สอบที่คลาดเคลื่อนที่มีคุณภาพด้านความตรงและความเที่ยงมากกว่าแบบสอบวินิจัยแบบเลือกตอบสามระดับและสองระดับ ตามลำดับ แต่แบบสอบแบบเลือกตอบสี่ระดับมีข้อจำกัดคือ ใช้เวลานาน และออกแบบข้อสอบให้มีคุณภาพได้ยาก ดังนั้นเพื่อเป็นการประหยัดเวลาและทรัพยากรในการทำแบบสอบ ผู้วิจัยจึงเลือกใช้แบบสอบแบบเลือกตอบสองระดับ โดยปรับให้มีการตอบในระดับคำตอบ และความมั่นใจในการตอบระดับคำตอบ ซึ่งเป็นการปรับปรุงแบบการตอบมาจากแบบสอบแบบเลือกตอบสี่ระดับ

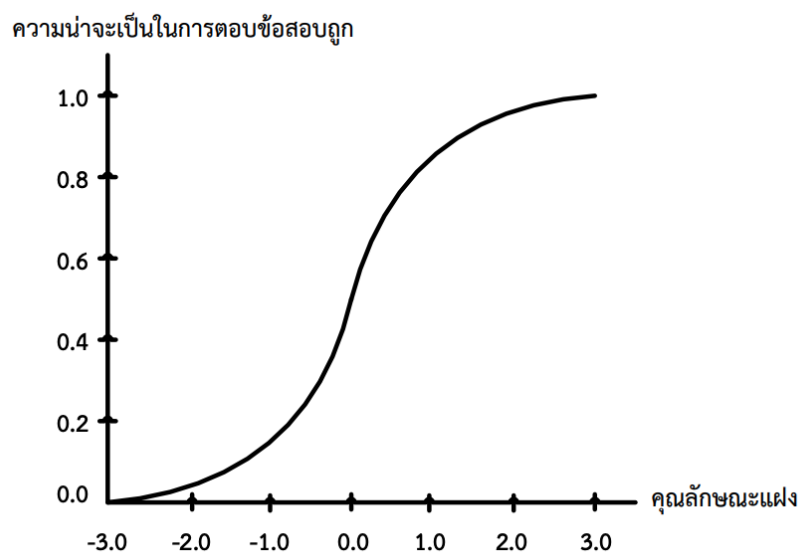
ตอนที่ 2 แนวคิดเกี่ยวกับทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) เป็นทฤษฎีการทดสอบแนวใหม่ซึ่งได้รับความนิยมและมีการนำไปใช้เพื่อพัฒนาคุณภาพของแบบวัดอย่างแพร่หลาย ในการนำเสนอแนวคิดเกี่ยวกับทฤษฎีการตอบสนองข้อสอบ แบ่งออกเป็น 2 หัวข้อ ดังนี้ 1) โมเดลการตอบสนองข้อสอบ และ 2) โมเดลการตอบสนองข้อสอบแบบตรวจให้คะแนนมากกว่า 2 ค่า

2.1 โมเดลการตอบสนองข้อสอบ

ทฤษฎีการตอบสนองข้อสอบ เข้ามามีบทบาทสำคัญในการอธิบายถึงความสามารถของผู้สอบด้วยแนวทางที่แตกต่างไปจากทฤษฎีการทดสอบแบบดั้งเดิม โดยมุ่งศึกษาคุณลักษณะแฝง (Latent Traits) ที่เป็นคุณลักษณะที่ซ่อนอยู่ภายในตัวบุคคล เช่น ความสามารถทางคณิตศาสตร์ ความสามารถทางภาษา ความสุข ความเศร้า เจตคติ โดยอาศัยโมเดลการตอบสนองข้อสอบ (Item Response Model) เข้ามาช่วยอธิบายความสัมพันธ์ระหว่างคุณลักษณะของข้อสอบและผู้สอบกับความน่าจะเป็น

ในการตอบข้อสอบถูกในรูปโค้งลักษณะข้อสอบ (Item Characteristic Curves: ICCs) หรือฟังก์ชันการตอบสนองข้อสอบ ซึ่งโดยทั่วไปนิยมเขียนอยู่ในรูปฟังก์ชันโลจิส (Logistic Function) (ศิริชัย กาญจนวาสี, 2555) ดังภาพ 3



ภาพ 3 โค้งลักษณะข้อสอบ

ที่มา: ปรับจาก Embetson & Reise, 2000, p.46 อ้างถึงใน สุชาติ สกลกิจรุ่งโรจน์, (ม.ป.ป.)

จากภาพ 3 จะเห็นได้ว่า แกนนอนแสดงถึงระดับคุณลักษณะแฝงของบุคคล ซึ่งมีค่าที่เป็นไปได้ตั้งแต่ $-\infty$ ถึง $+\infty$ แต่โดยปกติแล้วผลการวิเคราะห์ส่วนใหญ่จะให้ค่าตั้งแต่ -3 ถึง $+3$ ขณะที่แกนตั้งแสดงถึงค่าความน่าจะเป็นในการตอบข้อสอบถูกซึ่งมีค่าตั้งแต่ 0 ถึง 1 จากกราฟจะเห็นได้ว่า เมื่อระดับของคุณลักษณะแฝงเพิ่มขึ้น ความน่าจะเป็นในการตอบข้อสอบถูกจะเพิ่มขึ้นด้วย

เนื่องจากทฤษฎีการตอบสนองข้อสอบเป็นทฤษฎีที่ใช้ในการอธิบายความสัมพันธ์ระหว่างความสามารถที่มีอยู่ภายในบุคคลกับผลการตอบข้อสอบ ซึ่งเชื่อว่า พฤติกรรมการตอบสนองข้อสอบเป็นพฤติกรรมที่สังเกตได้ จะถูกกำหนดโดยคุณลักษณะหรือความสามารถที่มีอยู่ภายในบุคคลซึ่งเป็นพฤติกรรมที่สังเกตไม่ได้ ดังนั้นโมเดลการตอบสนองข้อสอบจึงเป็นโมเดลที่อธิบายความสัมพันธ์ระหว่างระดับความสามารถของบุคคล (θ) คุณลักษณะข้อสอบ ได้แก่ พารามิเตอร์ความยาก (b) อำนาจจำแนก (a) และโอกาสในการตอบข้อสอบได้ถูก (c)

โมเดลการตอบสนองข้อสอบ มีข้อตกลงเบื้องต้นสำคัญ 3 ประการคือ 1) ความเป็นเอกมิติ (Unidimensionality) หมายถึง ข้อสอบทุกข้อมุ่งวัดเพียงความสามารถเดียว 2) ความเป็นอิสระ (Local Independence) ระหว่างผู้สอบกับข้อสอบ แบ่งได้เป็น ความเป็นอิสระระหว่างผู้สอบคือ ผลการตอบข้อสอบข้อเดียวกันของผู้สอบแต่ละคนเป็นอิสระจากกัน และความเป็นอิสระระหว่างข้อสอบคือ ผลการตอบข้อสอบแต่ละข้อของผู้สอบคนเดียวกันเป็นอิสระจากกัน ด้วยคุณสมบัติดังกล่าว พารามิเตอร์ของผู้สอบและข้อสอบจึงไม่แปรเปลี่ยนไปตามกลุ่มผู้สอบหรือชุดข้อสอบที่นำมาใช้ ซึ่งแตกต่างจากทฤษฎีการทดสอบแบบดั้งเดิมที่มักให้ผลการวัดเปลี่ยนไปตามกลุ่มผู้สอบหรือชุดข้อสอบ และ 3) การเป็นฟังก์ชันต่อเนื่อง (Continuous Function) ที่สามารถอธิบายถึงความสัมพันธ์ระหว่างคุณลักษณะแฝงกับความน่าจะเป็นในการตอบข้อสอบได้ถูก (ศิริชัย กาญจนวาสี, 2555; Reckase, 2009)

การเลือกใช้โมเดลการตอบสนองข้อสอบจำเป็นต้องพิจารณาความเหมาะสมของโมเดลกับลักษณะอื่น ๆ ประกอบด้วย เช่น ลักษณะของข้อมูล มิติของตัวแปร จำนวนผู้สอบ จำนวนข้อสอบ/ข้อคำถาม เป็นต้น หากจำแนกโมเดลการตอบสนองข้อสอบตามลักษณะของข้อมูลสามารถจำแนกได้ 2 ประเภท (ศิริชัย กาญจนวาสี, 2555; De Ayala, 2009; Embretson & Reise, 2000; Hemberton, Swaminathan & Roger, 1991) ได้แก่ 1) โมเดลการตอบสนองข้อสอบแบบตรวจให้คะแนนสองค่า (Binary IRT Model หรือ Dichotomous IRT Model) สำหรับข้อมูลเมื่อตรวจให้คะแนนแล้วมีคะแนนที่เป็นไปได้สองค่า เป็น 1 กับ 0 เช่น ตอบถูกได้ 1 คะแนน ตอบผิดได้ 0 คะแนน ซึ่งอาจได้มาจากข้อสอบ/ข้อคำถามหลากหลายรูปแบบ เช่น แบบทดสอบแบบหลายตัวเลือกแต่มีคำตอบที่ถูกเพียงตัวเลือกเดียว แบบทดสอบที่ให้เลือกตอบแบบถูกผิด แบบประเมินบุคลิกภาพที่มีตัวเลือกเป็นจริงหรือไม่เป็นจริง แบบทดสอบเจตคติที่มีตัวเลือกเห็นด้วยหรือไม่เห็นด้วย มาตราประมาณค่าทางด้านพฤติกรรมที่มีตัวเลือกใช่หรือไม่ใช่ และ 2) โมเดลการตอบสนองข้อสอบแบบตรวจให้คะแนนมากกว่าสองค่า (Polytomous IRT Model) ใช้กับข้อสอบที่มีการตรวจให้คะแนนมากกว่า 2 ค่า เช่น ข้อสอบที่มีการตรวจให้คะแนนความรู้บางส่วน หรือข้อคำถามมาตราประมาณค่า

จากงานวิจัยที่เกี่ยวข้องกับโมเดลการตอบสนองข้อสอบของการให้คะแนนมากกว่าสองค่า โดยใช้ทฤษฎีการตอบสนองข้อสอบแบบตรวจให้คะแนนความรู้บางส่วนแบบทั่วไป (G-PCM) พบว่า ข้อสอบที่มีการให้คะแนนมากกว่า 2 ค่า จะให้สารสนเทศมากกว่าข้อสอบที่มีการให้คะแนน 2 ค่า (Dong, 1994) ดังนั้น ในงานวิจัยนี้จึงเลือกใช้โมเดลการตอบสนองข้อสอบแบบตรวจให้คะแนนมากกว่า 2 ค่า

2.2 โมเดลการตอบสนองข้อสอบแบบตรวจให้คะแนนมากกว่าสองค่า

โมเดลการตอบสนองข้อสอบแบบตรวจให้คะแนนมากกว่าสองค่า (Polytomous IRT Models) เป็นโมเดลการตอบสนองข้อสอบที่เหมาะสมกับแบบวัดตัวแปรด้านเจตพิสัยซึ่งข้อคำถามในแบบวัดมักมีลักษณะเป็นรายการคำตอบแบบเรียงลำดับหลายตัวเลือก (Ordered Categories Responses) และมีการตรวจให้คะแนนมากกว่าสองค่า โมเดลการตอบสนองข้อสอบในกลุ่มนี้โดยทั่วไปมี 6 โมเดลหลัก ดังนี้

1) Graded Response Model (GRM)

โมเดลนี้พัฒนาโดย Samejima ในปี ค.ศ. 1969 เป็นโมเดลการตอบสนองข้อสอบแบบตรวจให้คะแนนมากกว่าสองค่าที่เหมาะสมกับข้อมูลที่มีตัวเลือกรายการคำตอบแบบเรียงลำดับ เช่น มาตรฐานค่าของลิเคิร์ต (Likert Rating Scale) เป็นโมเดลที่ขยายมาจากโมเดลโลจิสแบบสองพารามิเตอร์ (Two-Parameter Logistic Model; 2PL) ซึ่งเป็นโมเดลการตอบสนองข้อสอบแบบตรวจให้คะแนนสองค่าที่มีแนวคิดที่ว่า ข้อสอบแต่ละข้อมีพารามิเตอร์ที่แตกต่างกัน 2 พารามิเตอร์ ซึ่งส่งผลต่อโอกาสหรือความน่าจะเป็นในการตอบข้อสอบถูกของผู้สอบ คุณสมบัติดังกล่าวคือ ความยากและอำนาจจำแนก ในการวิเคราะห์ข้อสอบจึงมีการนำค่าพารามิเตอร์ความยาก (Difficulty Parameter) และค่าพารามิเตอร์อำนาจจำแนก (Discrimination Parameter) มาใช้ในการคำนวณความน่าจะเป็นในการตอบข้อสอบถูกต้อง เนื่องจากเชื่อว่า โอกาสที่ผู้สอบจะตอบข้อสอบข้อใดข้อหนึ่งถูกต้องหรือไม่ขึ้นอยู่กับระดับความสามารถหรือคุณลักษณะแฝงของผู้สอบเพียงอย่างเดียว แต่ยังคงเกี่ยวข้องกับพารามิเตอร์ของข้อสอบแต่ละข้อด้วย

GRM เป็นโมเดลที่ขยายมาจากโมเดล 2PL จึงเป็นโมเดลที่เกี่ยวข้องกับ 2 พารามิเตอร์ โดยมีการเรียกชื่อพารามิเตอร์แตกต่างไปจากโมเดล 2PL คือ พารามิเตอร์ความชัน หรือ α_i (เทียบเท่ากับพารามิเตอร์อำนาจจำแนก) และพารามิเตอร์เทรชโฮลด์ หรือ β_{ij} (เทียบเท่ากับพารามิเตอร์ความยาก) การวิเคราะห์ GRM นั้น De Ayala (2009) ได้เสนอว่า ขนาดกลุ่มตัวอย่างที่ใช้ในการประมาณค่าพารามิเตอร์ควรมีไม่ต่ำกว่า 500 คน เพื่อให้สามารถวิเคราะห์ความสอดคล้องของโมเดล ประเมินความเป็นเอกมิติของคุณลักษณะแฝงที่ต้องการวัดได้ ทำให้สามารถประมาณค่าพารามิเตอร์ความชันได้ไม่ยากเกินไป และเพื่อให้มั่นใจได้ว่ามีข้อมูลการตอบในทุกตัวเลือกรายการคำตอบ อย่างไรก็ตาม ปรากฏการณ์นี้ไม่ใช่กฎตายตัว ในทางปฏิบัติการกำหนดขนาดกลุ่มตัวอย่างที่เหมาะสมควรพิจารณาจากปัจจัยหลากหลายด้าน เช่น วัตถุประสงค์ในการนำค่าพารามิเตอร์ที่ประมาณค่าได้ไปใช้ วิธีการประมาณค่าที่เลือกใช้ ความยาวของมาตราวัด ลักษณะการตอบข้อคำถามของผู้สอบ จำนวนข้อมูลที่ขาดหาย เป็นต้น

2) Modified Graded Response Model (M-GRM)

โมเดลนี้พัฒนาโดย Muraki (1990) เป็นโมเดลที่ปรับปรุงจาก GRM เหมาะสำหรับวิเคราะห์แบบสอบถามที่มีลักษณะเป็นมาตราประมาณค่า (Rating Scale) เช่น แบบวัดเจตคติ ข้อคำถามในมาตราวัดจะต้องมีจำนวนตัวเลือกรายการคำตอบเท่ากันแต่สามารถมีค่าพารามิเตอร์ความชันแตกต่างกันได้ M-GRM ใช้วิธีการคำนวณความน่าจะเป็นในการเลือกรายการคำตอบแบบสองขั้นตอนเหมือน GRM อย่างไรก็ตาม พารามิเตอร์เทรชโฮลด์ (β_{ij}) ใน M-GRM ถูกแบ่งออกเป็น 2 ส่วน ได้แก่ 1) พารามิเตอร์ตำแหน่ง (Location Parameter; b_i) ของข้อคำถามแต่ละข้อ และ 2) ชุดของพารามิเตอร์เทรชโฮลด์ (Threshold Parameter; c_j) ของมาตราวัดทั้งฉบับ ความแตกต่างระหว่าง GRM กับ M-GRM คือ GRM จะมีการประมาณค่าชุดของพารามิเตอร์เทรชโฮลด์ (β_{ij}) สำหรับข้อคำถามแต่ละข้อ แต่ M-GRM มีชุดของพารามิเตอร์เทรชโฮลด์ (c_j) เพียงชุดเดียวที่ใช้กับมาตราวัดทั้งฉบับ แต่มีการประมาณค่าพารามิเตอร์ตำแหน่ง (b_i) หนึ่งค่าสำหรับข้อคำถามแต่ละข้อ M-GRM ได้ชื่อว่าเป็น GRM แบบจำกัด (Restricted Model) เพราะโมเดลนี้มีจำนวนตัวเลือกรายการคำตอบ (Category Boundaries) ของข้อคำถามแต่ละข้อเท่ากันตลอดทั้งมาตราวัด ในขณะที่ GRM นั้น ข้อคำถามแต่ละข้ออาจมีจำนวนตัวเลือกรายการคำตอบของข้อคำถามแต่ละข้อแตกต่างกันได้ ดังนั้น M-GRM จึงมีการประมาณค่าพารามิเตอร์น้อยกว่า GRM จุดเด่นของ M-GRM เมื่อเปรียบเทียบกับ GRM คือ มีพารามิเตอร์ตำแหน่งแยกออกมาจากชุดของพารามิเตอร์เทรชโฮลด์ อย่างไรก็ตาม การเลือกใช้ขึ้นอยู่กับรูปแบบของข้อคำถามด้วย ถ้ามาตราวัดมีรูปแบบข้อคำถามหลากหลายแบบอยู่ในมาตราวัดฉบับเดียวกัน GRM จะมีความเหมาะสมมากกว่า M-GRM

3) Partial Credit Model (PCM)

PCM พัฒนาโดย Master (1982) เป็นโมเดลที่ถูกออกแบบมาสำหรับวิเคราะห์ข้อสอบที่มีกระบวนการตอบหลายลำดับขั้น โดยมีการตรวจให้คะแนนบางส่วน กล่าวคือ มีการให้คะแนนสำหรับแต่ละขั้นตอนที่ผู้สอบทำได้ ซึ่งพบได้บ่อยในแบบวัดผลสัมฤทธิ์ทางการเรียน เช่น โจทย์ปัญหาทางคณิตศาสตร์ เป็นต้น นอกจากนี้ PCM ยังสามารถนำมาใช้ในการวิเคราะห์มาตราวัดเจตคติหรือมาตราวัดบุคลิกภาพที่มีการประเมินความเชื่อหรือการตอบสนองต่อข้อความโดยมีการให้คะแนนแบบหลายค่า PCM จะแตกต่างจาก GRM และ M-GRM เพราะเป็นโมเดลแบบทางตรง (Direct Model) คือ คำนวณค่าความน่าจะเป็นในการเลือกรายการคำตอบของแต่ละตัวเลือกได้โดยตรง โมเดลนี้เป็นโมเดลที่มีการขยายต่อมาจากโมเดลโลจิสแบบหนึ่งพารามิเตอร์ (1PL) จึงมีคุณลักษณะมาตรฐานต่าง ๆ เช่นเดียวกับ Rasch Model เช่น การแยกจากกันของค่าพารามิเตอร์ข้อสอบกับค่าพารามิเตอร์บุคคล ใน PCM นั้น ข้อคำถามแต่ละข้อในมาตราวัดจะมีความชันเท่ากัน ดังนั้นพารามิเตอร์ความชันจึงไม่ปรากฏอยู่ในโมเดลนี้ คงเหลือเพียงพารามิเตอร์ความยากแบบลำดับขั้น (Step Difficulty; δ_{ij}) ที่ใช้อธิบายระดับความยากของการได้คะแนนจากระดับหนึ่งไปยังคะแนนอีก

ระดับหนึ่ง หรือจากตัวเลือกรายการคำตอบหนึ่งสู่อีกตัวเลือกหนึ่ง นอกจากนี้ ค่านี้อาจบ่งบอกถึงระดับคุณลักษณะแฝงบริเวณที่โค้งการเลือกรายการคำตอบ 2 รายการตัดกัน บางครั้งจึงนิยมเรียกว่า พารามิเตอร์จุดตัดรายการคำตอบ (Category Intersections Parameter) โดยข้อคำถามที่มีตัวเลือกรายการคำตอบ 5 รายการ จะมีพารามิเตอร์ความยากแบบลำดับขั้น 4 ค่า จุดเด่นของโมเดลนี้คือสามารถใช้คะแนนดิบรวมของมาตรวัดเพื่อประมาณค่าระดับคุณลักษณะแฝงได้ อย่างไรก็ตาม โมเดลนี้นำไปใช้ในทางปฏิบัติได้ยากเพราะในการประมาณค่าพารามิเตอร์อาจพบปัญหาว่าบริเวณจุดตัดโค้งการเลือกรายการคำตอบไม่ได้เรียงลำดับจากง่ายไปยากหรือจากระดับคุณลักษณะแฝงต่ำไปสูงทำให้เกิดปรากฏการณ์ที่เรียกว่า การผกผัน (Reversal) ซึ่งแสดงให้เห็นว่า มีตัวเลือกรายการคำตอบอย่างน้อยหนึ่งรายการที่เป็นตัวเลือกซึ่งไม่สอดคล้องกับคุณลักษณะแฝงที่ต้องการวัดหรือข้อมูลที่เก็บรวบรวมได้ไม่สอดคล้องกับโมเดล

4) Generalized Partial Credit Model (G-PCM)

G-PCM พัฒนาโดย Muraki (1992) เป็นโมเดลที่ปรับปรุงเพิ่มเติมจาก PCM ซึ่งกำหนดให้ข้อคำถามแต่ละข้อในมาตรวัดมีพารามิเตอร์ความชัน (α_i) แตกต่างกันได้ โดยพารามิเตอร์ความชันใน G-PCM เป็นค่าที่บ่งบอกถึงระดับการตอบสนองต่อตัวเลือกรายการคำตอบที่แตกต่างกันไปในแต่ละข้อคำถามซึ่งขึ้นอยู่กับระดับคุณลักษณะแฝงหรือระดับความสามารถ (θ) ที่เปลี่ยนแปลงไป ใน G-PCM นี้ หากข้อคำถามใดมีค่าพารามิเตอร์ความชันน้อยกว่า 1.0 โค้งการเลือกรายการคำตอบ (Category Response Curves) ที่ได้จะมีลักษณะแบนราบคล้ายใน PCM แต่หากค่าพารามิเตอร์ความชันมีค่ามากกว่า 1.0 โค้งการเลือกรายการคำตอบใน G-PCM จะมีลักษณะโด่งมากกว่าโค้งการเลือกรายการคำตอบที่พบใน PCM อย่างไรก็ตาม พารามิเตอร์จุดตัดรายการคำตอบ (δ_{ij}) ใน G-PCM มีความหมายเช่นเดียวกับ PCM

โมเดลนี้ถูกพัฒนาขึ้นโดย Muraki (1992) ซึ่งเป็นโมเดลที่พัฒนาต่อยอดมาจาก PCM ของ Master (1982) โมเดล PCM และ G-PCM มีพื้นฐานมาจากโมเดลการตอบสนองที่มี 1 พารามิเตอร์ จึงมีการวิเคราะห์คล้ายโมเดลของราสช์ (Rasch Model) และมีการคำนวณความน่าจะเป็นของการตอบแต่ละขั้นแบบขั้นตอนเดียว (Direct IRT Model) (Naumenko, 2014) โดย G-PCM เป็นโมเดลที่แก้ไขข้อจำกัดเรื่อง ความเป็นเอกมิติของเครื่องมือที่ใช้วัดคุณลักษณะ (เอมอร์ จังศิริพร ปรกรณ์, 2546) ข้อคำถามแต่ละข้อสามารถมีค่าพารามิเตอร์ความชันที่แตกต่างกัน ซึ่งความชันใน G-PCM เป็นค่าที่บ่งบอกระดับการตอบสนองต่อตัวเลือกรายการคำตอบที่แตกต่างกันในแต่ละข้อคำถามหรือค่าอำนาจจำแนก (Embetsen & Reise, 2000) ค่าอำนาจจำแนกใน G-PCM จะบ่งบอกถึงระดับการตอบในแต่ละลำดับค่าคะแนนที่เปลี่ยนไปในแต่ละข้อเมื่อระดับความสามารถ (θ) เปลี่ยนไป นอกจากนี้ G-PCM ยังมีความยืดหยุ่นสำหรับข้อสอบที่มีการเลือกตัวเลือกจำนวนมาก (Naumenko,

2014) ซึ่งความน่าจะเป็นที่จะตอบคำตอบลำดับคะแนนที่ k มากกว่า $k - 1$ สามารถแทนด้วยสมการ (Muraki, 1992) ดังนี้

$$C_{jk}(\theta) = P_{jk|k-1,k}(\theta) = \frac{P_{jk}(\theta)}{P_{j,k-1}(\theta) + P_{jk}(\theta)} = \frac{\exp[a_j(\theta - \delta_{jk})]}{1 + \exp[a_j(\theta - \delta_{jk})]}$$

$$P_{jk}(\theta) = \frac{C_{jk}}{1 - C_{jk}} P_{j,k-1,k}(\theta)$$

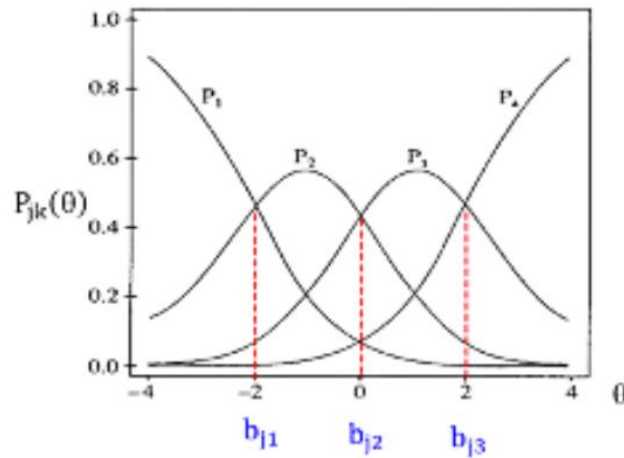
เมื่อ	k	คือ	ลำดับชั้นคะแนนที่ 1, 2, 3, ..., m_j
	$P_{jk}(\theta)$	คือ	ความน่าจะเป็นของคนที่มีความสามารถจะตอบได้คะแนน k
	$\frac{C_{jk}}{1 - C_{jk}}$	คือ	อัตราส่วนความน่าจะเป็นของสองเงื่อนไข ซึ่งอาจจะแสดงได้ในรูป $\exp[a_j(\theta - \delta_{jk})]$
	a_j	คือ	ค่าพารามิเตอร์อำนาจจำแนกของข้อ j
	b_{jk}	คือ	ค่าพารามิเตอร์ความยากของชั้นการตอบที่ k ในข้อ j เมื่อ $k = 1, 2, 3, \dots, m_j$

ค่าพารามิเตอร์ความยากของชั้นการตอบที่ k ในข้อ j (b_{jk}) เป็นจุดบนแกนระดับความสามารถของผู้สอบ (θ) ที่ฟังก์ชัน $P_{jk}(\theta)$ และ $P_{j,k-1}(\theta)$ ตัดกัน ซึ่งโค้งทั้ง 2 เป็นฟังก์ชันของการตอบใน 2 ลำดับกับค่าคะแนนในลำดับติดกันซึ่งตัดกันเพียงจุดเดียว ซึ่งถ้าจุดตัดต่างกันค่าพารามิเตอร์ความยากของชั้นการตอบก็จะแตกต่างกันและส่งผลต่อค่าความน่าจะเป็นของคนที่มีความสามารถจะตอบได้คะแนน k ก็จะต่างกัน ดังนี้

$$\text{ถ้า } \theta = b_{jk} \quad \text{จะได้ว่า } P_{jk}(\theta) = P_{j,k-1}(\theta)$$

$$\text{ถ้า } \theta > b_{jk} \quad \text{จะได้ว่า } P_{jk}(\theta) > P_{j,k-1}(\theta)$$

$$\text{ถ้า } \theta < b_{jk} \quad \text{จะได้ว่า } P_{jk}(\theta) < P_{j,k-1}(\theta)$$



ภาพ 4 โค้งรายการคำตอบ (CRC) จากการวิเคราะห์โดยใช้ G-PCM

ที่มา: Muraki (1992)

จากภาพ 4 แสดงโค้งรายการคำตอบ (Category Response Curves: CRC) ซึ่งมีค่าระดับความยากของขั้วการตอบ เท่ากับ b_{j1} , b_{j2} , b_{j3} เท่ากับ -2, 0 และ 2 ตามลำดับ แสดงให้เห็นว่า ผู้ที่มีระดับความสามารถ θ ณ จุดนั้นมีโอกาสที่จะตอบรายการคำตอบขั้นต่อไปสูงกว่ารายการคำตอบที่ผ่านมา เช่น ผู้ที่มีระดับความสามารถ -2 จะมีโอกาสที่จะตอบรายการคำตอบ P_2 มากกว่า P_1 นอกจากนี้ โค้งรายการคำตอบยังสามารถบอกได้ว่า ผู้ที่มีระดับความสามารถ -4 มีความน่าจะเป็นสูงสุดในการเลือกตอบรายการ P_1 ผู้ที่มีระดับความสามารถ -1 มีความน่าจะเป็นสูงสุดในการเลือกตอบรายการ P_2 ผู้ที่มีระดับความสามารถ 1 มีความน่าจะเป็นสูงสุดในการเลือกรายการ P_3 และผู้ที่มีระดับความสามารถ 4 มีความน่าจะเป็นสูงสุดในการเลือกตอบรายการ P_4

G-PCM ให้ค่าฟังก์ชันสารสนเทศคือ ฟังก์ชันสารสนเทศของแบบสอบ (Test Information) ซึ่งเป็นดัชนีที่บ่งชี้คุณภาพของแบบสอบทั้งฉบับ แสดงได้ดังสมการ

$$I(\theta) = \sum_{i=1}^k I_i(\theta)$$

เมื่อ $I(\theta)$ คือ ค่าฟังก์ชันสารสนเทศของแบบสอบ หรือค่าสารสนเทศที่ได้รับจากแบบสอบสำหรับผู้ที่มีความสามารถ θ

5) Rating Scale Model (RSM)

RSM เป็นกลุ่มของโมเดลที่มีหลากหลายรูปแบบและอาจก่อให้เกิดความสับสนได้ เพราะมีลักษณะคล้ายคลึงกับโมเดลอื่น ๆ เช่น PCM อย่างไรก็ตามแนวคิดของ Andrich (1978) มีความแตกต่างที่สำคัญคือ ข้อคำถามในมาตรวัดซึ่งมีรูปแบบการตอบแบบเดียวกันนั้น ข้อคำถามแต่ละข้อจะมีพารามิเตอร์ตำแหน่ง (Location Parameter; λ_i) เพียงค่าเดียว ซึ่งแสดงถึงความยากหรือง่ายของข้อคำถามนั้น ๆ ขณะที่พารามิเตอร์จุดตัดรายการคำตอบ (Category Intersections Parameter; δ_j) ใช้อธิบายช่วงเทรชโฮลด์ของแต่ละรายการคำตอบ (Category Threshold) ซึ่งมี J ช่วง ($J = K - 1$ โดย K คือ จำนวนตัวเลือกรายการคำตอบ) PCM กับ RSM แตกต่างกันที่ PCM ไม่มีแนวคิดเกี่ยวกับระดับความยากง่ายของแต่ละรายการคำตอบ (แต่ละขั้นตอนหรือแต่ละจุดตัดรายการคำตอบ) ในข้อคำถามนั้น ๆ ดังนั้น ข้อคำถามแต่ละข้อจึงมีระดับความยากของแต่ละขั้นตอน (Difficulties of the Steps) แตกต่างกันได้ ในขณะที่ RSM จะมีชุดของพารามิเตอร์จุดตัดรายการคำตอบเพียงชุดเดียวที่ใช้ตลอดทั้งมาตรวัด ตัวอย่างเช่น ข้อคำถามทางคณิตศาสตร์ที่ผู้สอบต้องแก้โจทย์ปัญหาหลายขั้นตอน ขั้นตอนต่าง ๆ ภายในข้อคำถามหรือระหว่างข้อคำถามอาจมีความยากง่ายต่างกันได้ อย่างไรก็ตามมาตรวัดเจตคติส่วนใหญ่ทุกข้อมีการตอบข้อคำถามรูปแบบเดียวกัน เช่น 1 หมายถึง ไม่เห็นด้วย 2 หมายถึง เห็นด้วยปานกลาง และ 3 หมายถึง เห็นด้วยอย่างยิ่ง จะสังเกตได้ว่า มาตรวัดนั้น ๆ ข้อคำถามแต่ละข้อมีระดับความยากง่ายหรือช่วงของแต่ละรายการคำตอบไม่แตกต่างกัน ความยากง่ายในแต่ละขั้นตอน (Step Difficulties หรือ Intersections) ซึ่งพบใน PCM สามารถแบ่งออกเป็น 2 ส่วนใน RSM ได้แก่ พารามิเตอร์ตำแหน่งคุณลักษณะแฝงที่ต้องการวัดของข้อคำถาม (Location Parameter; λ_i) และพารามิเตอร์จุดตัดรายการคำตอบ (Category Intersection Parameter; δ_j) RSM เหมาะกับข้อคำถามที่มีการกำหนดช่วงรายการคำตอบแบบคงที่ ซึ่งใช้กับข้อคำถามทุกข้อเหมือนกันตลอดมาตรวัด แต่ไม่เหมาะกับมาตรวัดที่มีลักษณะการตอบข้อคำถามที่หลากหลาย

6) Nominal Response Model (NRM)

NRM พัฒนาโดย Bock (1972) เป็นโมเดลสำหรับใช้วิเคราะห์ข้อสอบหรือข้อคำถามที่รายการคำตอบไม่ได้ถูกจัดเรียงลำดับ ได้แก่ ข้อสอบแบบหลายตัวเลือก (Multiple Choices) ข้อคำถามที่ใช้วัดเจตคติ ข้อคำถามสำหรับประเมินบุคลิกลักษณะ ความตั้งใจเริ่มแรกในการพัฒนาโมเดลนี้เกิดขึ้นเพื่อให้สามารถศึกษาคุณลักษณะของตัวลงซึ่งเป็นตัวเลือกหนึ่งในข้อสอบแบบหลายตัวเลือกด้วยเส้นร่องรอย (Trace Line) การคำนวณความน่าจะเป็นในการเลือกรายการคำตอบคำนวณได้จากพารามิเตอร์ที่เกี่ยวข้อง 2 พารามิเตอร์ ได้แก่ พารามิเตอร์ความชัน (Slope Parameter; α_{ix}) ซึ่งแสดงถึงความชันของเส้นร่องรอยหรือค่าอำนาจจำแนก และพารามิเตอร์จุดตัดรายการคำตอบ (Intercept Parameter; c_{ix})

ลักษณะของโมเดลการตอบสนองข้อสอบที่มีการตรวจให้คะแนนมากกว่าสองค่าทั้ง 6 โมเดลนี้ สามารถสรุปได้ ดังตาราง 3

ตาราง 3 สรุปลักษณะของโมเดลการตอบสนองข้อสอบที่มีการตรวจให้คะแนนมากกว่าสองค่า

โมเดล	ลักษณะรายการคำตอบ	จำนวนตัวเลือกในแต่ละข้อ		โมเดลพื้นฐาน	
		คำถามของแบบวัด		1PL	2PL
		ไม่ต้องเท่ากัน	ต้องเท่ากัน		
GRM	ตัวเลือกแบบเรียงลำดับ	✓			✓
M-GRM	ตัวเลือกแบบเรียงลำดับ		✓		✓
PCM	ตัวเลือกแบบเรียงลำดับ	✓		✓	
G-PCM	ตัวเลือกแบบเรียงลำดับ	✓			✓
RSM	ตัวเลือกแบบเรียงลำดับ		✓	✓	
NRM	ตัวเลือกแบบจำแนกกลุ่ม	✓			✓

ที่มา: สุชาติ สกลกิจรุ่งโรจน์ (ม.ป.ป.)

งานวิจัยนี้มีการตรวจให้คะแนนความรู้บางส่วน จึงเลือกใช้ Generalized Partial Credit Model (G-PCM) ในการวิเคราะห์ข้อมูล เนื่องจากมีความยืดหยุ่นมากกว่า Partial Credit Model (PCM) และมีข้อตกลงเบื้องต้นที่สอดคล้องกับธรรมชาติของข้อสอบ นั่นคือ ในการวิเคราะห์ข้อมูลของข้อคำถามแต่ละข้อโดยใช้ G-PCM นั้น สามารถมีค่าพารามิเตอร์ความชันหรือค่าอำนาจจำแนกแตกต่างกันได้ ซึ่งจะสอดคล้องกับความเป็นจริงมากกว่า

ตอนที่ 3 แนวคิดเกี่ยวกับวิธีการตรวจให้คะแนนความรู้บางส่วน

การวัดและประเมินผลทางการศึกษา ปัจจุบันนิยมใช้ข้อสอบเลือกตอบแบบหลายตัวเลือก (Multiple Choice) ที่มีตัวเลือกเพียงตัวเดียว (วัลลี สัตยาศัย, ม.ป.ป.) โดยมีการให้คะแนนแบบตอบถูกได้ 1 คะแนน และตอบผิดได้ 0 คะแนน ซึ่งการวัดประเมินผลโดยใช้ข้อสอบเลือกตอบแบบหลายตัวเลือกมีข้อจำกัดคือ ผู้เรียนมีโอกาสในการเดาคำตอบได้ ซึ่งจะทำให้คะแนนที่ได้ไม่สะท้อนความสามารถที่แท้จริงของผู้เรียนได้ (เอมอร์ จังศิริพรปรกรณ์, 2545) ดังนั้น จึงไม่สามารถระบุได้ว่านักเรียนมีความรู้จริง หรือไม่มีความรู้ หรือมีความรู้บางส่วน (Chan et al., 2007) จากข้อจำกัดดังกล่าวจึงเป็นที่มาของการศึกษาเกี่ยวกับการให้คะแนนความรู้บางส่วน

มโนทัศน์เกี่ยวกับวิธีการให้คะแนนความรู้บางส่วน

การให้คะแนนความรู้บางส่วนคือ การให้คะแนนที่ใกล้เคียงกับความสามารถที่แท้จริงของผู้ตอบ โดยวิธีการให้คะแนนความรู้บางส่วนจะช่วยลดโอกาสในการเดาของผู้ตอบ (P. N. K. Lau et al., 2011) และส่งผลต่อความเที่ยงและความตรงในการวัด (Frary, 1980) ซึ่งการประเมินความรู้ของผู้ตอบส่วนใหญ่ใช้แบบสอบแบบหลายตัวเลือก (Multiple Choice) เนื่องจากแบบสอบแบบหลายตัวเลือกมีความน่าเชื่อถือและสามารถใช้ในการประเมินผู้เรียนจำนวนมากได้ สามารถวัดความรู้ความสามารถได้กว้างขวาง มีความเป็นปรนัยในการตรวจให้คะแนน ใช้เวลาน้อยในการตรวจให้คะแนน สามารถนำข้อมูลไปวิเคราะห์เพื่อวินิจฉัยปัญหาการเรียนรู้ของผู้เรียนได้ (เอมอร์ จังศิริพรปกรณ์, 2545) แต่แบบสอบแบบหลายตัวเลือกมีข้อจำกัดคือ ผู้สอบมีโอกาสในการเดา ทำให้ผลการวัดไม่เป็นไปตามความสามารถที่แท้จริงของผู้สอบ และไม่สามารถให้สารสนเทศที่เพียงพอที่จะจำแนกผู้เรียนที่มีความสามารถต่าง ๆ ได้ เช่น ผู้เรียนที่มีความรู้เต็มที่ (Full Knowledge) มีความรู้บางส่วน (Partial Knowledge) จนถึงมีความรู้ที่ผิดเต็มที่ (Full Misconception) (เอมอร์ จังศิริพรปกรณ์, 2545) จากข้อจำกัดดังกล่าว Lord & Novick (1968) (อ้างถึงใน เอมอร์ จังศิริพรปกรณ์, 2545) ได้เสนอวิธีการแก้ไขข้อจำกัดของแบบสอบแบบหลายตัวเลือก 4 วิธี ดังนี้

1) การให้น้ำหนักคะแนนแก่ข้อสอบที่แตกต่างกัน (Differential Item Weighting)

การให้น้ำหนักคะแนนแก่ข้อสอบที่แตกต่างกันคือ การให้น้ำหนักคะแนนมากสำหรับข้อสอบที่มีคุณภาพสูง และให้น้ำหนักคะแนนน้อยสำหรับข้อสอบที่มีคุณภาพต่ำ โดยมีเกณฑ์พิจารณาคุณภาพข้อสอบคือ ค่าความยาก ค่าอำนาจจำแนก ค่าความตรง หรือการพิจารณาจากผู้เชี่ยวชาญการให้คะแนนแก่ข้อสอบควรใช้ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory) ซึ่งเป็นการประมาณค่าตามระดับความสามารถของผู้สอบ จะทำให้น้ำหนักของคะแนนมีค่าคงที่ไม่แปรเปลี่ยนไปตามกลุ่มตัวอย่าง แต่ถ้าใช้ทฤษฎีการวัดแบบดั้งเดิม (Classical Theory) จะทำให้น้ำหนักของคะแนนแต่ละข้อมีค่าไม่คงที่และแปรเปลี่ยนไปตามกลุ่มตัวอย่าง (Stanley & Wang, 1970)

2) การให้น้ำหนักคะแนนแก่ตัวเลือกที่แตกต่างกัน (Differential Option Weighting)

การให้น้ำหนักคะแนนแก่ตัวเลือกที่แตกต่างกัน มีหลักการการให้คะแนน 2 วิธี (Smith, 1987) ดังนี้

2.1) วิธีการตรวจให้คะแนนของเดวิสและฟิฟเฟอร์ คือการให้น้ำหนักคะแนนแก่ตัวเลือกจากวิจารณ์ของผู้เชี่ยวชาญหรือจากทฤษฎีของโครงสร้างความรู้ ซึ่งวิธีนี้จะช่วยให้คะแนนที่ได้มีความเที่ยงสูง แต่มีความยุ่งยากในการปฏิบัติมาก

2.2) *วิธีการตรวจให้คะแนนของแฟรริย์* คือการให้น้ำหนักคะแนนแก่ตัวเลือกจากข้อมูลเชิงประจักษ์ เช่น คะแนนมาตรฐานโดยเฉลี่ยของแต่ละตัวเลือกที่ผู้ตอบเลือก หรือความสัมพันธ์ระหว่างตัวเลือกที่ผู้ตอบเลือกกับคะแนนรวมทั้งหมด วิธีนี้เหมาะกับการให้คะแนนความรู้บางส่วน และช่วยให้ค่าความตรงของการวัดสูงขึ้น

3) การเปลี่ยนแปลงโครงสร้างของข้อสอบ (Changing the Item Structure)

การเปลี่ยนแปลงโครงสร้างของข้อสอบ อาจปรับเปลี่ยนตัวเลือกให้มีคำตอบถูกมากกว่า 1 ตัวเลือก หรือไม่มีคำตอบถูกเลย และให้นักเรียนระบุจำนวนตัวเลือกที่ถูก ในบางกรณี อาจจะมีการหักคะแนนในการเลือกตัวเลือกที่ผิด โดยมีจุดมุ่งหมายเพื่อลดโอกาสในการเดาและประเมินความรู้บางส่วนของผู้ตอบ วิธีการให้คะแนนที่สำคัญ มีดังนี้

3.1) *แบบสอบเลือกตอบแบบถูก-ผิด (Multiple True-False)* จะมีคำตอบที่ถูกต้องหลายคำตอบหรือไม่มีเลยก็ได้ โดยผู้ตอบจะต้องระบุว่าตัวเลือกใดถูกหรือตัวเลือกใดผิด ซึ่งแบบสอบแบบนี้มีความเที่ยงและความตรงสูงกว่าแบบสอบเลือกตอบแบบธรรมดาที่มีคำตอบถูก 1 ตัวเลือก (Fisbie & Woods, 1974)

3.2) *แบบสอบเลือกตอบแบบหลายคำตอบ (Multiple Answer)* เป็นแบบสอบที่มีคำตอบถูกมากกว่า 1 ตัวเลือก โดยผู้ตอบจะต้องเลือกตัวเลือกที่เป็นคำตอบที่ถูกให้ได้ทุกตัว คะแนนที่ได้เท่ากับผลรวมของตัวเลือกที่ตอบถูก และจะมีการหักคะแนนการตอบตัวเลือกผิด เพื่อลดโอกาสในการเดา และเพื่อประเมินความรู้บางส่วน ซึ่งแบบสอบเลือกตอบแบบหลายคำตอบมีข้อจำกัดคือ มีความยุ่งยากในการตรวจให้คะแนน และคะแนนรวมในแต่ละข้อไม่เท่ากันเนื่องจากคำตอบที่ถูกต้องในแต่ละข้อไม่เท่ากัน ดังนั้น วิธีนี้จึงไม่นิยมที่จะนำมาใช้ในการพิจารณาความรู้บางส่วน

3.3) *แบบสอบเลือกตอบชนิดเชิงซ้อน (Complex Multiple Choice)* เป็นการนำตัวเลือกของแบบสอบเลือกตอบแบบหลายคำตอบมาจัดกลุ่มใหม่ โดยมีชุดตัวเลือกที่ถูกต้องทั้งหมดอยู่ในกลุ่มเดียวกัน และเป็นคำตอบที่ถูกต้องเพียงกลุ่มเดียวเท่านั้น ซึ่งถ้าจัดกลุ่มไม่เหมาะสมจะทำให้เป็นการแนะนำคำตอบที่ถูกต้องได้ และจะทำให้ผู้ตอบที่มีความรู้เพียงบางส่วนตอบถูกได้ ดังนั้น วิธีนี้จึงไม่นิยมที่จะนำมาใช้ในการพิจารณาความรู้บางส่วน

3.4) *แบบสอบเลือกตอบชนิดจับคู่ (Simple or Multiple Matching Test)* เป็นรูปแบบหนึ่งของข้อสอบหลายตัวเลือก ซึ่งจะใช้ในการกรณข้อสอบมีตัวเลือกที่เหมือนกัน ประกอบด้วย ชุดข้อความ เรียกว่า ตัวยืน (Premises) และชุดของตัวเลือก เรียกว่า คำตอบ (Response) (Waugh, 2013) โดยผู้ตอบต้องเลือกคำตอบที่มีความสัมพันธ์กับคำถามหรือตัวยืน แบบสอบเลือกตอบชนิดจับคู่ แบ่งได้เป็น 2 ประเภท ได้แก่ 3.4.1) การจับคู่อย่างง่าย (Simple Matching) คือ แบบสอบที่มีจำนวนคำถามเท่ากับจำนวนคำตอบ และ 3.4.2) การจับคู่แบบพหุ (Multiple Matching) คือ แบบสอบที่มีจำนวนคำตอบมากกว่าจำนวนคำถาม ซึ่งจะช่วยลดโอกาสในการเดาได้

มากกว่าการจับคู่อย่างง่าย และแบบสอบการจับคู่แบบพหุจะช่วยให้ค่าความตรงและความเที่ยงสูงกว่าแบบสอบเลือกตอบธรรมดา

4) การเปลี่ยนแปลงวิธีการตอบ (Changing the Response Method)

การเปลี่ยนแปลงวิธีการตอบเป็นวิธีการประเมินตนเองที่ดีที่สุด ทำให้ได้สารสนเทศเกี่ยวกับความรู้ของผู้สอบในแต่ละข้อ ผู้สอบจะเป็นกำหนดน้ำหนักคะแนนของการตอบด้วยตนเอง ซึ่งวิธีนี้จะช่วยลดโอกาสในการเดาและทำให้ได้สารสนเทศที่สะท้อนความสามารถที่แท้จริงของผู้สอบ การเปลี่ยนแปลงวิธีการตอบแบ่งเป็น 2 วิธี ดังนี้

4.1) *วิธีการปรับคะแนนเพื่อแก้การเดาโดยการลงโทษ* โดยปกติถ้าผู้สอบตอบผิดจะได้คะแนนเป็น 0 แต่วิธีการนี้ ถ้าผู้สอบตอบผิดอันเนื่องมาจากการเดาจะได้คะแนนติดลบ โดยมีสูตรในการคำนวณ ดังนี้

$$X_c = \frac{R - W}{K - 1}$$

เมื่อ	X_c	คือ	คะแนนที่ปรับแก้แล้ว
	R	คือ	จำนวนข้อสอบที่ตอบถูก
	W	คือ	จำนวนข้อสอบที่ตอบผิด
	K	คือ	จำนวนตัวเลือกของข้อสอบแต่ละข้อ

วิธีการแก้การเดาโดยการลงโทษเหมาะที่จะใช้กับแบบสอบความเร็ว (Speed Test) มากกว่าแบบสอบไม่จำกัดเวลา เนื่องจากการทำแบบสอบไม่จำกัดเวลาผู้สอบอาจมีความรู้บางส่วนในการตัดตัวเลือกออกได้ ดังนั้น การใช้วิธีการแก้การเดาโดยการลงโทษจึงไม่เหมาะสมที่จะพิจารณาการมีความรู้บางส่วน

4.2) *วิธีการให้รางวัลแก่ข้อที่ไม่ตอบ* โดย Ebel (1965) ได้เสนอวิธีสำหรับการให้รางวัลแก่ข้อที่ไม่ตอบ ดังนี้

$$X_c = R + \frac{O}{K}$$

เมื่อ	X_c	คือ	คะแนนที่ปรับแก้แล้ว
	R	คือ	จำนวนข้อที่ตอบถูก
	O	คือ	จำนวนข้อที่ไม่ตอบ
	K	คือ	จำนวนตัวเลือกของข้อสอบแต่ละข้อ

วิธีการให้รางวัลแก่ข้อที่ไม่ตอบมีข้อจำกัดคือ ข้อที่ผู้สอบเว้นไว้หรือไม่ตอบนั้นไม่สามารถระบุได้ว่ามาจากการที่ผู้สอบไม่มีความรู้หรือผู้สอบทำไม่ทัน

5) วิธีการปรับวิธีการตอบ

วิธีการปรับวิธีการตอบ เป็นวิธีการที่ทำให้ผู้สอบแสดงความรู้ออกมาให้มากที่สุดเพื่อเป็นสารสนเทศในการพิจารณาระดับความรู้ของผู้สอบ ซึ่งมีนักการศึกษาได้ศึกษาเกี่ยวกับวิธีการปรับวิธีการตอบหลายวิธี ดังนี้

5.1) *วิธีตัดตัวลวงเลือกตัวถูก (The Number Right Elimination Testing: NRET)* โดย Lau, Lau, Hong & Usop (2014) เสนอวิธีนี้ โดยให้ผู้สอบตัดตัวลวงที่แน่ใจว่าไม่ใช่คำตอบออก เลือก 1 ตัวเลือกที่เป็นคำตอบที่ถูกต้อง โดยตัวลวงที่ไม่แน่ใจให้ตอบ “ไม่แน่ใจ” ดังนั้น ถ้าตัดตัวลวงถูก 1 ตัวจะได้ 1 คะแนน ตัวลวงทั้งหมดจะมี $k - 1$ ตัว เมื่อ k เป็นจำนวนตัวเลือก ถ้าตัดตัวถูกโดยเข้าใจว่าเป็นตัวลวงจะถูกหักคะแนน $-(k - 1)$ เมื่อ k เป็นจำนวนตัวเลือก แต่ถ้าไม่แน่ใจให้เลือกตอบ “ไม่แน่ใจ” จะได้ 0 คะแนน การที่มีตัวเลือกไม่แน่ใจ เพื่อเป็นการเพิ่มตัวเลือกในการพิจารณาเลือกคำตอบและหลีกเลี่ยงการเดาของนักเรียน นอกจากนี้ วิธีตัดตัวลวงเลือกตัวถูกยังเป็นรูปแบบที่ส่งเสริมให้นักเรียนได้ฝึกกลยุทธ์ในการทำข้อสอบ และส่งเสริมให้นักเรียนแก้ปัญหาเพื่อตอบข้อที่ถูกต้อง นอกจากนี้ วิธีนี้ยังสามารถวัดความเข้าใจที่คลาดเคลื่อนได้ และสามารถจำแนกผู้เรียนที่มีความรู้สมบูรณ์ มีความรู้บางส่วน ไม่มีความรู้ มีความรู้คลาดเคลื่อนบางส่วน และมีความรู้คลาดเคลื่อนทั้งหมดได้

การให้คะแนน

- | | |
|--|-------------------------|
| (1) เลือกคำตอบที่ถูกต้อง | ได้ 1 คะแนน |
| (2) ตัดตัวลวงถูกต้อง 1 ตัว | ได้ 1 คะแนน |
| (3) ตัดตัวถูกออกโดยเข้าใจว่าเป็นตัวลวง | ถูกหัก $-(k - 1)$ คะแนน |
| (4) ตอบ “ไม่แน่ใจ” | ได้ 0 คะแนน |

การแปลความหมายคะแนน เมื่อข้อสอบมี 4 ตัวเลือก

- | | |
|-----------------------|---|
| (1) ได้ 4 คะแนน | หมายถึง ผู้สอบมีความรู้สมบูรณ์ |
| (2) ได้ 1, 2, 3 คะแนน | หมายถึง ผู้สอบมีความรู้บางส่วน |
| (3) ได้ 0 คะแนน | หมายถึง ผู้สอบไม่มีความรู้ |
| (4) ได้ -1, -2 คะแนน | หมายถึง ผู้สอบมีความรู้คลาดเคลื่อนบางส่วน |
| (5) ได้ -3 คะแนน | หมายถึง ผู้สอบมีความรู้คลาดเคลื่อนทั้งหมด |

5.2) *วิธีการเลือกชุดย่อยของคำตอบ (The Subset Selection Method: SST)* เป็นวิธีของ Dressel & Schmidt (1953 อ้างถึงใน Simon et al., 1997) ซึ่งผู้สอบจะต้องเลือกชุดคำตอบย่อยที่มีคำตอบถูกอยู่ในชุดนั้น

การให้คะแนนขึ้นอยู่กับจำนวนตัวเลือกที่มีอยู่ในชุดย่อยของคำตอบนั้น

(1) ถ้าชุดคำตอบย่อยไม่มีคำตอบที่ถูกต้องอยู่ คะแนนจะถูกหักเท่ากับจำนวนตัวเลือกที่เลือก

(2) ถ้าไม่เลือกตัวเลือกใดเลย จะได้ 0 คะแนน

5.3) *วิธีการทดสอบความน่าจะเป็น (Probability Testing)* เป็นการรายงานความน่าจะเป็นของตัวเลือกที่ผู้สอบคิดว่าเป็นคำตอบถูก มีช่วงของการให้คะแนนระหว่าง 0-1 โดยแบ่งระดับความรู้เป็น 5 ประเภท (Wallsten, Budescu & Zwick, 1993 อ้างถึงใน เอมอร์ จังศิริพรภรณ์, 2545) ดังนี้

(1) มีความรู้เต็มที่ (Full Knowledge) คือ ให้ความน่าจะเป็นเท่ากับ 1 แก่คำตอบที่ถูกต้อง

(2) มีความรู้บางส่วน (Partial Knowledge) คือ ให้ความน่าจะเป็นมากกว่า 0 แต่ไม่ถึง 1 ($0 < p < 1$) แก่คำตอบที่ถูกต้อง

(3) มีความรู้ที่ผิดบางส่วน (Partial Misinformation) คือ ให้ความน่าจะเป็นมากกว่า 0 แต่ไม่ถึง 1 ($0 < p < 1$) แก่ตัวเลือกที่ผิด

(4) มีความรู้ที่ผิดเต็มที่ (Full Misinformation) คือ ให้ความน่าจะเป็นเท่ากับ 0 แก่คำตอบที่ถูกต้อง

(5) ไม่มีความรู้ (Absence of Knowledge) คือ ให้ความน่าจะเป็นแก่ทุกตัวเลือกเท่ากันหรือไม่ตอบ

5.4) *วิธีการแสดงความมั่นใจ (Confidence Marking)* วิธีนี้ผู้สอบจะต้องแสดงความมั่นใจในคำตอบที่คิดว่าถูกโดยใช้ C-Point Scale หรืออาจจะระบุว่า ไม่แน่ใจ ค่อนข้างแน่ใจ หรือแน่ใจมาก ซึ่งมีการแบ่งระดับความรู้ออกเป็น 5 ระดับ (Dressel & Schmidt, 1953 อ้างถึงใน เอมอร์ จังศิริพรภรณ์, 2545) ดังนี้

(1) Full Knowledge คือ การเลือกคำตอบได้ถูกต้องด้วยความมั่นใจสูงสุด

(2) Partial Knowledge คือ การเลือกคำตอบได้ถูกต้องด้วยความมั่นใจระดับต่ำ

(3) Partial Misinformation คือ การเลือกคำตอบผิดด้วยความมั่นใจระดับต่ำ

(4) Full Misinformation คือ การเลือกคำตอบผิดด้วยความมั่นใจระดับสูง

(5) Absence of Knowledge คือ การไม่เลือกคำตอบหรือเว้นว่างไว้

5.5) *วิธีการจัดอันดับอย่างสมบูรณ์ (Complete Ordering)* เป็นวิธีที่ผู้สอบจะต้องจัดอันดับตัวเลือก ซึ่งสามารถจำแนกความรู้ได้เป็น 3 ระดับ (อ้างถึงใน เอมอร์ จังศิริพรภรณ์, 2545) ดังนี้

(1) Full Knowledge คือ การจัดอันดับให้คำตอบถูกอยู่ในตำแหน่งสูงสุด

(2) Partial Knowledge คือ การจัดอันดับให้คำตอบถูกอยู่ในตำแหน่งกลาง

(3) Absence of Knowledge คือ การจัดอันดับให้คำตอบถูกอยู่ในตำแหน่งต่ำสุด

5.6) วิธีการจัดอันดับเป็นบางส่วน (Partial Ordering) เป็นวิธีที่ผสมระหว่างวิธีการตัดตัวลง และวิธีการจัดอันดับอย่างสมบูรณ์ ผู้สอบจะต้องจัดอันดับตัวเลือกที่ไม่สามารถตัดออกไปได้ ซึ่งสามารถจำแนกความรู้ได้เป็น 5 ระดับ (De Finetti, 1965 อ้างถึงใน Simon et al., 1997) ดังนี้

(1) Full Knowledge คือ การจัดอันดับแก่คำตอบถูกตำแหน่งสูงสุด

(2) Partial Knowledge คือ การจัดอันดับแก่หลายตัวเลือก โดยคำตอบจะถูกจัดอันดับในตำแหน่งสูงสุด

(3) Partial Misinformation คือ การจัดอันดับแก่คำตอบถูกในตำแหน่งที่ต่ำกว่าตัวลง

(4) Full Misinformation คือ การจัดอันดับแก่ตัวลง โดยตัดคำตอบถูกออก

(5) Absence of Knowledge คือ การจัดอันดับแก่ตัวถูกในตำแหน่งที่ต่ำที่สุดหรือไม่ตอบ

ตอนที่ 4 แนวคิดเกี่ยวกับสมรรถนะทางวิทยาศาสตร์

การประเมินด้านวิทยาศาสตร์ของ PISA ให้ความสำคัญกับการประยุกต์ใช้ความรู้ทางวิทยาศาสตร์ในชีวิตประจำวัน โดย PISA 2025 จะเน้นการประเมินว่า นักเรียนสามารถแสดงสมรรถนะทางวิทยาศาสตร์ 3 ด้านได้อย่างเหมาะสมในบริบทต่าง ๆ ทั้งในระดับส่วนบุคคล ระดับท้องถิ่น/ประเทศ และระดับโลก ซึ่งการประเมินดังกล่าวจะแตกต่างจากการเรียนวิทยาศาสตร์ในโรงเรียนโดยทั่วไปที่มักมุ่งเน้นความรู้ด้านเนื้อหา แต่กรอบการประเมินนี้เป็นมุมมองวิทยาศาสตร์ในมุมมองที่กว้างขึ้นเพื่อให้เห็นถึงความรู้ประเภทต่าง ๆ ที่จำเป็นต่อการมีส่วนร่วมของสมาชิกในสังคมร่วมสมัย ตามจุดประสงค์ของการประเมิน PISA 2025 สามารถจำแนกองค์ประกอบที่เกี่ยวข้องกับการศึกษาด้านวิทยาศาสตร์ออกเป็น 4 ด้านที่สัมพันธ์กัน ดังนี้

1) บริบท (Context) หมายถึง สถานการณ์หรือประเด็นปัญหาในระดับส่วนบุคคล ระดับท้องถิ่น/ประเทศ และระดับโลก ทั้งที่เป็นเรื่องในปัจจุบันหรือในอดีตที่ผ่านมาซึ่งจำเป็นต้องมีความเข้าใจเรื่องวิทยาศาสตร์และเทคโนโลยี

2) ความรู้ทางวิทยาศาสตร์ (Scientific Knowledge) หมายถึง ความเข้าใจในข้อเท็จจริง แนวคิดหลัก และทฤษฎีสำคัญ ที่ทำให้เกิดความรู้พื้นฐานทางวิทยาศาสตร์ โดยความรู้ทางวิทยาศาสตร์ ประกอบด้วย

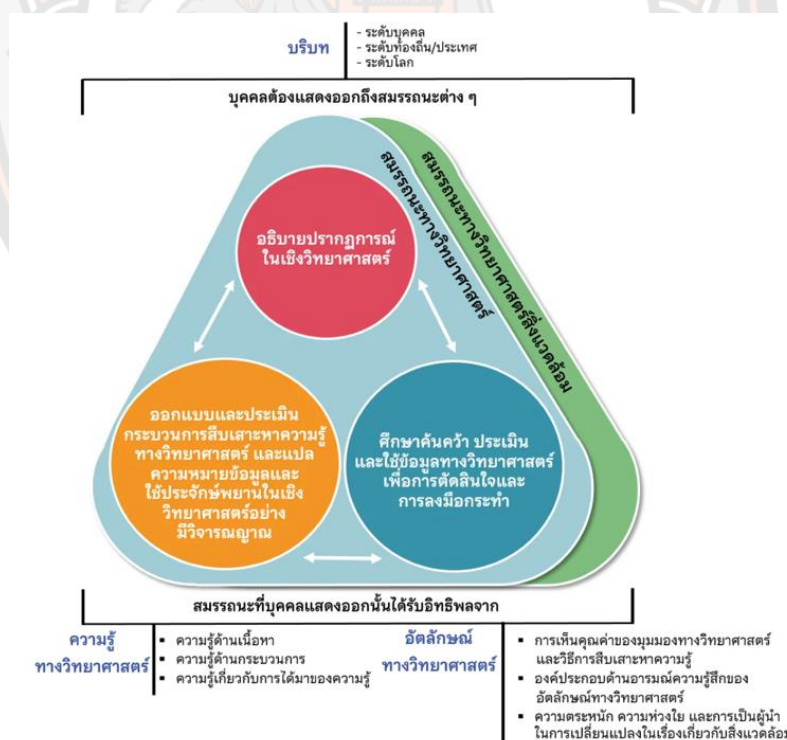
2.1) ความรู้ด้านเนื้อหา (Content Knowledge) เป็นความรู้เกี่ยวกับข้อเท็จจริง แนวคิด และทฤษฎีเกี่ยวกับโลกธรรมชาติที่วิทยาศาสตร์ได้สร้างขึ้น

2.2) ความรู้ด้านกระบวนการ (Procedural Knowledge) เป็นความรู้เกี่ยวกับวิธีการในการสร้างแนวคิดต่าง ๆ

2.3) ความรู้เกี่ยวกับการได้มาของความรู้ (Epistemic Knowledge) เป็นความเข้าใจในเหตุผลพื้นฐานของกระบวนการสร้างความรู้

3) สมรรถนะทางวิทยาศาสตร์ (Science Competencies) หมายถึง ความสามารถในการอธิบายปรากฏการณ์ในเชิงวิทยาศาสตร์ การออกแบบและประเมินกระบวนการสืบเสาะหาความรู้ทางวิทยาศาสตร์ และการแปลความหมายข้อมูลและใช้ประจักษ์พยานในเชิงวิทยาศาสตร์อย่างมีวิจารณญาณ และการศึกษาค้นคว้า ประเมิน และใช้ข้อมูลทางวิทยาศาสตร์เพื่อการตัดสินใจและการลงมือกระทำ

4) อัตลักษณ์ทางวิทยาศาสตร์ (Science Identity) หมายถึง กลุ่มของลักษณะอุปนิสัย การเป็นผู้นำในการเปลี่ยนแปลง เจตคติต่อวิทยาศาสตร์ และต้นทุนทางวิทยาศาสตร์ของบุคคล ซึ่งสามารถบ่งชี้ได้จากความสนใจในวิทยาศาสตร์และเทคโนโลยี การให้คุณค่ากับวิธีการสำรวจตรวจสอบทางวิทยาศาสตร์ด้วยวิธีการที่เหมาะสม และการรับรู้และตระหนักถึงปัญหาสิ่งแวดล้อม



ภาพ 5 องค์ประกอบของกรอบการประเมินด้านวิทยาศาสตร์

ที่มา: สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี (2560)

4.1 ความหมายของสมรรถนะทางวิทยาศาสตร์

PISA 2025 ให้นิยาม สมรรถนะทางวิทยาศาสตร์ ไว้ดังนี้ สมรรถนะทางวิทยาศาสตร์ (Science Competency) หมายถึง ความสามารถของบุคคลที่มีส่วนร่วมในการอภิปรายเกี่ยวกับวิทยาศาสตร์ ความยั่งยืน และเทคโนโลยี ได้อย่างสมเหตุสมผล เพื่อนำไปสู่การลงมือกระทำได้ บริบทที่ใช้ในการประเมิน

PISA ให้ความสำคัญในการประเมินคือ การใช้ความรู้ทางวิทยาศาสตร์ในบริบทหรือสถานการณ์ต่าง ๆ อย่างหลากหลายในการจัดการกับประเด็นทางวิทยาศาสตร์และสิ่งแวดล้อม โดยสถานการณ์ที่ใช้จะไม่ขึ้นกับบริบทของประเทศใดประเทศหนึ่ง แต่บริบทนั้นจะต้องสามารถแสดงหรือสะท้อนสมรรถนะทางวิทยาศาสตร์ของผู้เรียนในบริบททั้งสามระดับ ได้แก่ บริบทระดับส่วนบุคคล (เช่น ประเด็นที่เกิดกับตัวเอง ครอบครัว หรือเพื่อน) บริบทระดับสังคมหรือบริบทระดับท้องถิ่น/ประเทศ (เช่น ประเด็นที่ส่งผลกระทบต่อสังคม วัฒนธรรม สุขภาพ หรือชีวิตมนุษย์) และบริบทระดับโลก (เช่น ประเด็นทางวิทยาศาสตร์ที่เป็นข่าวในสื่อ หรือมีผลกระทบต่อสิ่งแวดล้อมหรือต่อโลกอนาคต)

คำถามของการประเมิน PISA จึงอยู่ในสถานการณ์ที่เป็นส่วนหนึ่งในชีวิตจริงของนักเรียน และไม่จำกัดอยู่เฉพาะสถานการณ์ในโรงเรียนเท่านั้น แต่จะเป็นสถานการณ์ที่อาจเกี่ยวข้องกับตัวนักเรียนเอง ครอบครัว ชุมชน หรือสถานการณ์ของโลกก็ได้ แม้ว่าประเด็นเกี่ยวกับเทคโนโลยีและสิ่งแวดล้อมมักเป็นบริบททั่วไปในการประเมิน แต่คำถามที่อยู่ในบริบทของประวัติศาสตร์ที่เหมาะสมก็สามารถนำมาใช้ประเมินความเข้าใจเกี่ยวกับกระบวนการและความก้าวหน้าของความรู้ทางวิทยาศาสตร์ได้

ตาราง 4 บริบทที่ใช้ในการประเมิน 3 ระดับ กับตัวอย่างสถานการณ์ที่เป็นส่วนหนึ่งในชีวิตประจำวัน

สถานการณ์	บริบท		
	ระดับส่วนบุคคล	ระดับท้องถิ่น/ประเทศ	ระดับโลก
สุขภาพและโรคภัย	การดูแลรักษาสุขภาพ อุบัติเหตุ โภชนาการ การฉีควัคซีน	การควบคุมโรค การ แพร่เชื้อในสังคม การ เลือกอาหาร โรคอ้วน สุขภาพชุมชน	การระบาดของโรคที่ แพร่กระจายไปทั่วโลก ความมั่นคงทางอาหาร วิถีชีวิตที่ดีต่อสุขภาพ
ทรัพยากรธรรมชาติ	การใช้วัสดุต่าง ๆ ส่วน บุคคล ประเภทของ อาหารและพลังงาน	การควบคุมขนาด ประชากรมนุษย์ คุณภาพชีวิต ความ	แหล่งพลังงาน หมุนเวียนและพลังงาน ที่ใช้แล้วหมดไป ระบบ

สถานการณ์	บริบท		
	ระดับส่วนบุคคล	ระดับท้องถิ่น/ประเทศ	ระดับโลก
	การบริโภคอาหารในท้องถิ่น การเลือกผลิตภัณฑ์ที่ไม่ผสมนมจากสัตว์ และการกินมังสวิรัต	มั่นคง การผลิตและการกระจายอาหาร การจัดหาพลังงาน ผลกระทบต่อสิ่งแวดล้อมจากการทำเหมืองแร่และอุตสาหกรรม การผลิตพลังงานจากแหล่งพลังงานหมุนเวียน	ในธรรมชาติ การเติบโตของประชากร การใช้ประโยชน์จากสิ่งมีชีวิตชนิดต่าง ๆ และที่ดินอย่างยั่งยืน ความหลากหลายทางชีวภาพและคุณค่าของ ความหลากหลายทางชีวภาพ
ผลกระทบต่อสิ่งแวดล้อมและการเปลี่ยนแปลงภูมิอากาศ	การมีพฤติกรรมที่ยั่งยืนในการรีไซเคิล และการลดการใช้ทรัพยากร	การกระจายของประชากร การจัดการกับขยะ ผลกระทบต่อสิ่งแวดล้อม การใช้ระบบเกษตรกรรมฟื้นฟู	ความยั่งยืนทางสิ่งแวดล้อม การจัดการมลพิษและคุณภาพอากาศ การสูญเสียดิน/มวลชีวภาพ การสูญเสียพันธุ์ รั้งใหญ่ ปราบปรามการทะเลาะ
ภัยอันตราย	การประเมินความเสี่ยงในการเลือกดำเนินชีวิต	การประเมินความเสี่ยงจากการเปลี่ยนแปลงกะทันหัน (เช่น แผ่นดินไหว สภาพอากาศเลวร้าย) การเปลี่ยนแปลงอย่างช้า ๆ และต่อเนื่อง (เช่น การกัดเซาะชายฝั่ง การเกิดตะกอน) การจดจำ	ภัยคุกคามที่เกิดจากการเปลี่ยนแปลงภูมิอากาศ ผลกระทบจากการติดต่อสื่อสารยุคใหม่ พลังงานและการผลิตพลังงาน (เช่น การผลิตปิโตรเลียม ด้วยเทคโนโลยีขุดเจาะชั้นหิน พลังงานนิวเคลียร์ พลังงาน

สถานการณ์	ระดับส่วนบุคคล	บริบท	
		ระดับท้องถิ่น/ประเทศ ในหน้า	ระดับโลก จากแก๊สธรรมชาติ)
ความก้าวหน้าและ ความท้าทายของ วิทยาศาสตร์และ เทคโนโลยีร่วมสมัย	แง่มุมทางวิทยาศาสตร์ เกี่ยวกับเทคโนโลยี ใหม่ ๆ (เช่น การ ปรับแต่งยีน เสมือนจริง)	วัสดุ เครื่องมือ และ กระบวนการใหม่ การ ตัดแปรรูปนวัตกรรม เทคโนโลยีเกี่ยวกับ สุขภาพ การคมนาคม ขนส่ง การใช้ ปัญญาประดิษฐ์	การสำรวจอวกาศ จุด กำเนิด และโครงสร้าง ของเอกภพ

ที่มา: สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี (2560)

4.2 องค์ประกอบของสมรรถนะทางวิทยาศาสตร์

PISA ให้นิยาม การรู้เรื่องวิทยาศาสตร์ (Scientific Literacy) ไว้ว่า เป็นความสามารถในการเชื่อมโยงสิ่งต่าง ๆ เข้ากับประเด็นที่เกี่ยวข้องกับวิทยาศาสตร์ และแนวคิดทางวิทยาศาสตร์ได้อย่างไตร่ตรอง บุคคลที่รู้เรื่องวิทยาศาสตร์ (Scientific Literacy Person) จะสื่อสารพูดคุยในประเด็นที่เกี่ยวข้องกับวิทยาศาสตร์และเทคโนโลยีอย่างเป็นเหตุเป็นผลจำเป็นต้องใช้สมรรถนะที่กำหนดองค์ประกอบตาม PISA (2017) ซึ่งแบ่งออกเป็น 3 องค์ประกอบ แต่ละองค์ประกอบจะประกอบด้วย 5 ตัวบ่งชี้ (สำนักวิชาการและมาตรฐานทางการศึกษา สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน, 2558) ดังนี้

องค์ประกอบที่ 1 การอธิบายปรากฏการณ์ในเชิงวิทยาศาสตร์ (Explain Phenomena Scientifically)

เป็นสมรรถนะที่จำเป็นสำหรับการรู้เรื่องวิทยาศาสตร์ การแสดงออกถึงสมรรถนะนี้ บุคคลที่รู้เรื่องวิทยาศาสตร์ต้องสามารถระลึกถึงความรู้ด้านเนื้อหาที่เหมาะสมในสถานการณ์ที่กำหนดให้ และใช้ความรู้เพื่อแปลความหมายและให้คำอธิบายต่อปรากฏการณ์ต่าง ๆ รวมถึงการวาดแบบจำลองทางวิทยาศาสตร์เพื่อใช้อธิบายปรากฏการณ์ที่เกิดขึ้นในชีวิตประจำวัน การบรรยาย และการตีความปรากฏการณ์ การคาดการณ์หรือการพยากรณ์การเปลี่ยนแปลงที่อาจจะเกิดขึ้น รวมถึงการระบุ ว่า คำบรรยาย คำอธิบายใดสมเหตุสมผลหรือไม่อย่างไร คำคาดการณ์จะเป็นไปได้หรือไม่ด้วยเหตุผลอะไร เป็นต้น

โดยสรุป ผู้ที่มีสมรรถนะการอธิบายปรากฏการณ์ในเชิงวิทยาศาสตร์ ควรแสดงความสามารถ ดังนี้

- 1.1 ใช้องค์ความรู้ทางวิทยาศาสตร์อธิบายปรากฏการณ์ได้อย่างเหมาะสม
- 1.2 เลือกแบบจำลองทางวิทยาศาสตร์มาใช้เป็นตัวแทนของสิ่งของประกอบคำอธิบายปรากฏการณ์
- 1.3 พยากรณ์การเปลี่ยนแปลงที่อาจเกิดขึ้นได้จริงจากการตีความปรากฏการณ์
- 1.4 เสนอสมมติฐานที่ใช้ในการอธิบายปรากฏการณ์
- 1.5 อธิบายถึงความเป็นไปได้ในการนำองค์ความรู้ทางวิทยาศาสตร์ไปใช้ประโยชน์กับสังคม

องค์ประกอบที่ 2 การประเมินและออกแบบกระบวนการสืบเสาะหาความรู้ทางวิทยาศาสตร์ (Evaluate and Design Scientific Enquiry)

บุคคลที่รู้เรื่องวิทยาศาสตร์ต้องมีความสามารถในการประเมินและออกแบบการสืบเสาะหาความรู้ทางวิทยาศาสตร์ เพื่อใช้ในการสร้างความรู้ที่เชื่อถือได้เกี่ยวกับโลกและธรรมชาติ การแสดงออกถึงสมรรถนะด้านนี้ บุคคลต้องสามารถประเมินข้อค้นพบทางวิทยาศาสตร์อย่างมีวิจารณญาณ แยกแยะคำถามทางวิทยาศาสตร์ว่าคำถามใดสามารถตอบได้ด้วยการสำรวจทางวิทยาศาสตร์ สมรรถนะนี้จำเป็นต้องใช้ความรู้เกี่ยวกับลักษณะสำคัญของการสำรวจตรวจสอบทางวิทยาศาสตร์ เช่น การทดสอบที่เที่ยงตรงต้องทำอย่างไร ต้องเปรียบเทียบกับอะไร ควรคุมตัวแปรใด และเปลี่ยนแปลงตัวแปรใด ต้องค้นคว้าสารและข้อมูลอะไรเพิ่มเติม และต้องทำอย่างไรจึงจะเก็บข้อมูลที่ต้องการได้ นอกจากนี้ยังต้องรู้ถึงความสำคัญและคุณค่าของงานวิจัยต่าง ๆ ในแง่มุมที่อาจมีความคลุมเครือ การสรุปไม่สมเหตุสมผล ไม่มีข้อมูลมากพอหรือมีความลำเอียงได้ เป็นต้น

โดยสรุป ผู้ที่มีสมรรถนะการประเมินและออกแบบกระบวนการสืบเสาะหาความรู้ทางวิทยาศาสตร์ ควรแสดงความสามารถดังนี้

- 2.1 ระบุปัญหาที่ต้องการสำรวจในการศึกษาทางวิทยาศาสตร์
- 2.2 บอกได้ว่าประเด็นปัญหาหรือคำถามใดสามารถตรวจสอบได้ด้วยวิธีการทางวิทยาศาสตร์
- 2.3 เสนอวิธีสำรวจตรวจสอบปัญหาทางวิทยาศาสตร์
- 2.4 ประเมินวิธีสำรวจตรวจสอบปัญหาทางวิทยาศาสตร์
- 2.5 ประเมินและอธิบายวิธีการต่าง ๆ ที่นักวิทยาศาสตร์ใช้ยืนยันความน่าเชื่อถือของข้อมูล

รวมถึงคำอธิบายปรากฏการณ์ที่ปราศจากความคิดเห็นส่วนตัว

องค์ประกอบที่ 3 การแปลความหมายข้อมูลและการใช้ประจักษ์พยานในเชิงวิทยาศาสตร์ (Interpret Data and Evidence Scientifically)

บุคคลที่มีสมรรถนะการแปลความหมายข้อมูลและใช้ประจักษ์พยานในเชิงวิทยาศาสตร์ต้องแสดงออกถึงความสามารถในการตีความข้อมูลและหลักฐานทางวิทยาศาสตร์ที่ใช้ในการสร้างคำกล่าว

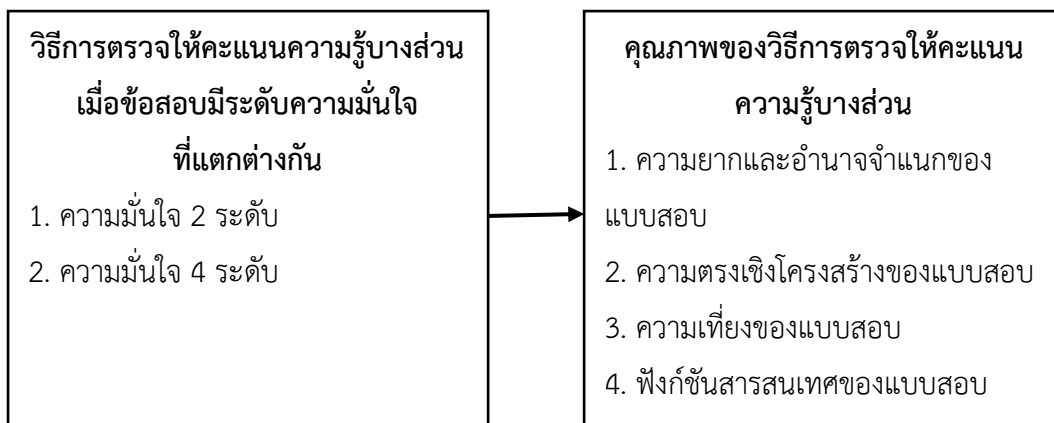
อ้างหรือลงข้อสรุป นำเสนอข้อมูลที่ได้รับในรูปแบบอื่น ๆ เช่น ใช้คำพูดของตนเอง แบนภาพ หรือการ แสดงอื่น ๆ แทนได้ ซึ่งสมรรถนะนี้จำเป็นต้องใช้เครื่องมือทางคณิตศาสตร์ในการวิเคราะห์หรือสรุป ข้อมูลและใช้ความสามารถในการใช้วิธีการพื้นฐานของประจักษ์พยาน ข้อมูล หรือประเมินข้อสรุปที่ ผู้อื่นสร้างขึ้นว่าสอดคล้องกับประจักษ์พยานที่มีหรือไม่ รวมถึงสามารถโต้แย้งอย่างสมเหตุสมผล

โดยสรุป ผู้ที่มีสมรรถนะการแปลความหมายข้อมูลและการใช้ประจักษ์พยานในเชิง วิทยาศาสตร์ ควรแสดงความสามารถดังนี้

- 3.1 แปลงข้อมูลจากรูปแบบการนำเสนอเดิมไปสู่รูปแบบอื่น
- 3.2 วิเคราะห์และตีความหมายข้อมูล และลงข้อสรุปอย่างสมเหตุสมผล
- 3.3 ระบุข้อสันนิษฐาน ประจักษ์พยาน (หลักฐาน) และเหตุผลในเรื่องที่เกี่ยวกับวิทยาศาสตร์
- 3.4 บอกความแตกต่างระหว่างข้อโต้แย้งที่อยู่บนพื้นฐานของประจักษ์พยานทางวิทยาศาสตร์ กับข้อโต้แย้งที่ใช้ฐานแนวคิดอื่น ๆ
- 3.5 ประเมินข้อโต้แย้งทางวิทยาศาสตร์และประจักษ์พยานจากแหล่งที่มาที่หลากหลาย เช่น หนังสือพิมพ์ อินเทอร์เน็ต และวารสาร

ตอนที่ 5 กรอบแนวคิดการวิจัย

การวิจัยครั้งนี้ ผู้วิจัยได้ศึกษาคุณภาพของวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบ มีระดับความมั่นใจที่ต่างกันคือ ความมั่นใจ 2 ระดับ และความมั่นใจ 4 ระดับ เนื่องจากความ มั่นใจ เป็นการพิจารณาคุณลักษณะภายในของแต่ละบุคคลอย่างหนึ่ง ซึ่งในการศึกษาทางวิทยาศาสตร์ ยังพบงานวิจัยไม่มากนักที่นำมาตรปรมาณความมั่นใจมาใช้ในการพัฒนาแบบสอบ โดยคุณภาพของ วิธีการตรวจให้คะแนนความรู้บางส่วนที่ผู้วิจัยต้องการศึกษามี 4 ด้าน คือ ความยากและอำนาจ จำแนกของแบบสอบ ความตรงเชิงโครงสร้างของแบบสอบ ความเที่ยงของแบบสอบ และฟังก์ชัน สารสนเทศของแบบสอบ กรอบแนวคิดในการวิจัยเป็นดังภาพ 6



ภาพ 6 กรอบแนวคิดในการวิจัย



บทที่ 3

วิธีดำเนินงานวิจัย

การวิจัยเรื่อง การเปรียบเทียบคุณภาพของแบบสอบเลือกตอบสองระดับด้านสมรรถนะทางวิทยาศาสตร์ที่มีระดับความมั่นใจที่ต่างกัน โดยประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบแบบตรวจให้คะแนนความรู้บางส่วนแบบทั่วไป มีวัตถุประสงค์ 4 ประการ ได้แก่ 1) เพื่อเปรียบเทียบความยากและอำนาจจำแนกของแบบสอบ 2) เพื่อเปรียบเทียบความตรงเชิงโครงสร้างของแบบสอบ 3) เพื่อเปรียบเทียบความเที่ยงของแบบสอบ และ 4) เพื่อเปรียบเทียบฟังก์ชันสารสนเทศของแบบสอบ โดยมีขั้นตอนในการดำเนินการวิจัย ดังนี้

ประชากรและกลุ่มตัวอย่าง

ประชากร

ประชากรที่ใช้ในการวิจัย ได้แก่ นักเรียนระดับชั้นมัธยมศึกษาปีที่ 4-6 จำนวน 21,663 คน จำนวน 57 โรงเรียน สังกัดสำนักงานเขตพื้นที่การศึกษามัธยมศึกษาพิษณุโลก อุตรดิตถ์ (สำนักงานเขตพื้นที่การศึกษามัธยมศึกษาพิษณุโลก อุตรดิตถ์, 2566) รายละเอียดดังตาราง 5

ตาราง 5 ประชากรจำแนกตามจังหวัด จำนวนโรงเรียน จำนวนห้องเรียน และจำนวนนักเรียน

จังหวัด	จำนวนโรงเรียน	จำนวนห้องเรียน	จำนวนนักเรียน
พิษณุโลก	39	475	15,066
อุตรดิตถ์	18	218	6,597
รวม	57	693	21,663

กลุ่มตัวอย่าง

กลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้ได้จากการสุ่มแบบหลายขั้นตอน (Multi-stage Sampling) จากนักเรียนระดับชั้นมัธยมศึกษาปีที่ 4-6 สังกัดสำนักงานเขตพื้นที่การศึกษามัธยมศึกษาพิษณุโลก อุตรดิตถ์ จำนวน 39 โรงเรียนในเขตจังหวัดพิษณุโลก ได้กลุ่มตัวอย่างจำนวน 750 คน ขั้นตอนในการเก็บรวบรวมข้อมูล มีรายละเอียดดังนี้

ขั้นตอนที่ 1 กำหนดขนาดกลุ่มตัวอย่าง ซึ่งงานวิจัยนี้เป็นการเปรียบเทียบคุณภาพของแบบสอบ 2 ฉบับ โดยประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบแบบตรวจให้คะแนนความรู้บางส่วนแบบ

ทั่วไป ในการวิเคราะห์พารามิเตอร์ต่าง ๆ ตามทฤษฎีการตอบสนองข้อสอบแบบตรวจให้คะแนนมากกว่า 2 ค่า (Polytomous IRT) สามารถใช้ขนาดกลุ่มตัวอย่างขั้นต่ำแต่ละกลุ่มเท่ากับ 375 คนได้ (De Ayala, 1994) ซึ่งงานวิจัยนี้มีการเก็บรวบรวมข้อมูลโดยแบ่งกลุ่มตัวอย่างเป็น 2 กลุ่ม ได้แก่ กลุ่มที่ 1 ผู้สอบจะต้องตอบแบบสอบที่มีระดับความมั่นใจ 2 ระดับ และกลุ่มที่ 2 ผู้สอบจะต้องตอบแบบสอบที่มีระดับความมั่นใจ 4 ระดับ ดังนั้น งานวิจัยนี้ต้องเก็บรวบรวมข้อมูลจากกลุ่มตัวอย่างวิจัยทั้ง 2 กลุ่ม กลุ่มละไม่ต่ำกว่า 375 คน รวมทั้งหมดเป็นจำนวน 750 คน จำนวนกลุ่มตัวอย่างจำแนกตามอำเภอในจังหวัดพิษณุโลก ซึ่งเป็นโรงเรียนในสังกัดสำนักงานเขตพื้นที่การศึกษามัธยมศึกษาพิษณุโลก อุตรดิตถ์

ขั้นตอนที่ 2 สุ่มโรงเรียนจากโรงเรียนในจังหวัดพิษณุโลก สังกัดสำนักงานเขตพื้นที่การศึกษามัธยมศึกษาพิษณุโลก อุตรดิตถ์ โดยวิธีการสุ่มอย่างง่าย เพื่อเก็บข้อมูลกลุ่มตัวอย่างจำนวน 750 คน โดยมีเงื่อนไขในการเก็บข้อมูลคือ นักเรียนต้องเป็นนักเรียนในระดับชั้นมัธยมศึกษาตอนปลาย จากนั้นแบ่งกลุ่มตัวอย่างเป็น 2 กลุ่ม ดังนี้

กลุ่มที่ 1 ผู้สอบจะต้องตอบแบบสอบที่มีระดับความมั่นใจ 2 ระดับ ได้แก่ มั่นใจ และไม่มั่นใจ

กลุ่มที่ 2 ผู้สอบจะต้องตอบแบบสอบที่มีระดับความมั่นใจ 4 ระดับ ได้แก่ มั่นใจมาก มั่นใจ ไม่มั่นใจ และเดา

งานวิจัยนี้ผู้วิจัยกำหนดขนาดกลุ่มตัวอย่างวิจัยไว้จำนวน 750 คน แบ่งเป็นกลุ่มตัวอย่างกลุ่มที่ 1 จำนวน 375 คน และกลุ่มที่ 2 จำนวน 375 คน ซึ่งแต่ละกลุ่มมีขนาดกลุ่มตัวอย่างเพียงพอที่จะวิเคราะห์ผลตามทฤษฎีการตอบสนองข้อสอบแบบตรวจให้คะแนนมากกว่า 2 ค่า เนื่องจากในการวิเคราะห์ผลตามทฤษฎีดังกล่าวสามารถใช้ขนาดกลุ่มตัวอย่างขั้นต่ำแต่ละกลุ่มเท่ากับ 375 คนได้ (De Ayala, 1994) จำนวนกลุ่มตัวอย่างจำแนกตามอำเภอในจังหวัดพิษณุโลก ได้ข้อมูลดังตาราง 6 และจำนวนกลุ่มตัวอย่างจำแนกตามโรงเรียนในจังหวัดพิษณุโลก ได้ข้อมูลดังตาราง 7

ตาราง 6 จำนวนกลุ่มตัวอย่างจำแนกตามอำเภอในจังหวัดพิษณุโลก

อำเภอ	กลุ่มที่ 1		กลุ่มที่ 2	
	จำนวนกลุ่มตัวอย่าง (คน)	ร้อยละ	จำนวนกลุ่มตัวอย่าง (คน)	ร้อยละ
เมืองพิษณุโลก	75	20.00	120	32.00
บางระกำ	80	21.33	116	30.93
พรหมพิราม	29	7.73	-	-
บางกระทุ่ม	-	-	43	11.47

อำเภอ	กลุ่มที่ 1		กลุ่มที่ 2	
	จำนวนกลุ่ม ตัวอย่าง (คน)	ร้อยละ	จำนวนกลุ่ม ตัวอย่าง (คน)	ร้อยละ
วังทอง	107	28.53	48	12.80
ชาติตระการ	-	-	14	3.73
วัดโบสถ์	40	10.67	-	-
เนินมะปราง	-	-	14	3.73
นครไทย	44	11.73	20	5.33
รวม	375	100	375	100

ตาราง 7 จำนวนกลุ่มตัวอย่างจำแนกตามโรงเรียนในจังหวัดพิษณุโลก

โรงเรียน	กลุ่มที่ 1		กลุ่มที่ 2	
	จำนวนกลุ่ม ตัวอย่าง (คน)	ร้อยละ	จำนวนกลุ่ม ตัวอย่าง (คน)	ร้อยละ
อำเภอ เมืองพิษณุโลก				
1. โรงเรียนจ่านกร้อง	12	3.20	-	-
2. โรงเรียนเฉลิมขวัญสตรี	-	-	18	4.80
3. โรงเรียนดอนทองวิทยา	27	7.20	-	-
4. โรงเรียนเตรียมอุดมศึกษา ภาคเหนือ	7	1.87	-	-
5. โรงเรียนท่าทองพิทยาคม	-	-	36	9.60
6. โรงเรียนบ้านกร่างพิทยาคม	11	2.93	-	-
7. โรงเรียนพิษณุโลกพิทยาคม	-	-	43	11.47
8. โรงเรียนพุทธชินราชพิทยา	18	4.80	-	-
9. โรงเรียนวังน้ำคู้ศึกษา	-	-	23	6.13
10. โรงเรียนวิทยาศาสตร์จุฬาภรณราชวิทยาลัย พิษณุโลก	-	-	-	-
รวม	75	20.00	120	32.00
อำเภอ บางระกำ				
11. โรงเรียนชุมแสงสงคราม (อุตรคอนารักษ์อุปถัมภ์)	26	6.93	-	-

โรงเรียน	กลุ่มที่ 1		กลุ่มที่ 2	
	จำนวนกลุ่ม ตัวอย่าง (คน)	ร้อยละ	จำนวนกลุ่ม ตัวอย่าง (คน)	ร้อยละ
12. โรงเรียนบางระกำวิทยศึกษาศาสตร์	54	14.40	-	-
13. โรงเรียนประชาสงเคราะห์วิทยา	-	-	116	30.93
รวม	80	21.33	116	30.93
อำเภอ พรหมพิราม				
14. โรงเรียนดงประคำพิทยาคม	-	-	-	-
15. โรงเรียนพรหมพิรามวิทยา	14	3.73	-	-
16. โรงเรียนวังมะด่างพิทยาคม	15	4.00	-	-
รวม	29	7.73	-	-
อำเภอ บางกระทุ่ม				
17. โรงเรียนบางกระทุ่มพิทยาคม	-	-	43	11.47
18. โรงเรียนเนินกุ่มวิทยา	-	-	-	-
รวม	-	-	43	11.47
อำเภอ วังทอง				
19. โรงเรียนทรัพย์ไพรวัลย์วิทยา	9	2.40	-	-
20. โรงเรียนน้ำรินพิทยาคม	-	-	22	5.87
21. โรงเรียนเนินสะอาดวิทยา	15	4.00	-	-
22. โรงเรียนบ้านกลางพิทยาคม	-	-	26	6.93
23. โรงเรียนวังทองพิทยาคม	35	9.33	-	-
24. โรงเรียนวังพิถุสวิทยศึกษาศาสตร์	6	1.60	-	-
25. โรงเรียนสฤชดีเสนาพิทยาคม	-	-	-	-
26. โรงเรียนหนองพระพิทยา	42	11.20	-	-
รวม	107	28.53	48	12.80
อำเภอ ชาติตระการ				
27. โรงเรียนชาติตระการ	-	-	14	3.73
28. โรงเรียนสวนเมี่ยงวิทยา	-	-	-	-
รวม	-	-	14	3.73
อำเภอ วัดโบสถ์				

โรงเรียน	กลุ่มที่ 1		กลุ่มที่ 2	
	จำนวนกลุ่ม ตัวอย่าง (คน)	ร้อยละ	จำนวนกลุ่ม ตัวอย่าง (คน)	ร้อยละ
29. โรงเรียนคันโช้งพิทยาคม	8	2.13	-	-
30. โรงเรียนวัดโบสถ์ศึกษา	32	8.53	-	-
รวม	40	10.67	-	-
อำเภอ เนินมะปราง				
31. โรงเรียนไทรย้อยพิทยาคม	-	-	-	-
32. โรงเรียนเนินมะปรางศึกษาวิทยา	-	-	14	3.73
33. โรงเรียนวังโพรงพิทยาคม	-	-	-	-
รวม	-	-	14	3.73
อำเภอ นครไทย				
34. โรงเรียนนครชุมพิทยา รัชมิ่งคลาภิเษก	-	-	3	0.80
35. โรงเรียนนครไทย	44	11.73	-	-
36. โรงเรียนนครบางยางพิทยาคม	-	-	-	-
37. โรงเรียนนาบัววิทยา	-	-	6	1.60
38. โรงเรียนบ่อโพธิ์วิทยา	-	-	9	2.40
39. โรงเรียนยางโกลนวิทยา	-	-	2	0.53
รวม	44	11.73	20	5.33
รวมทั้งหมด	375	100	375	100

เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัยครั้งนี้ ได้แก่ แบบสอบเลือกตอบด้านสมรรถนะทางวิทยาศาสตร์ รายวิชาวิทยาศาสตร์ชีวภาพ ระดับชั้นมัธยมศึกษาตอนปลาย จำนวน 2 ฉบับ ฉบับละ 20 ข้อ ซึ่งมีลักษณะเป็นแบบสอบเลือกตอบสองระดับ โดยในแต่ละข้อคำถาม ประกอบด้วยคำถาม 2 ส่วนคือ คำถามระดับแรกหรือระดับเนื้อหา (Content Tier) วัดความรู้เนื้อหาสำคัญด้วยแบบสอบเลือกตอบ 4 ตัวเลือก และคำถามระดับสองหรือระดับความมั่นใจ (Confidence Tier) วัดความมั่นใจของคำตอบในระดับเนื้อหา โดยมีระดับความมั่นใจที่แตกต่างกันในแต่ละฉบับ ดังนี้

แบบสอบเลือกตอบด้านสมรรถนะทางวิทยาศาสตร์ ฉบับที่ 1 ที่มีระดับความมั่นใจ 2 ระดับ ได้แก่ มั่นใจ และ ไม่มั่นใจ

แบบสอบเลือกตอบด้านสมรรถนะทางวิทยาศาสตร์ ฉบับที่ 2 ที่มีระดับความมั่นใจ 4 ระดับ ได้แก่ มั่นใจมาก มั่นใจ ไม่มั่นใจ และ เคา

แบบสอบเลือกตอบสองระดับด้านสมรรถนะทางวิทยาศาสตร์ รายวิชาวิทยาศาสตร์ชีวภาพ ระดับชั้นมัธยมศึกษาตอนปลาย ที่มีระดับความมั่นใจที่แตกต่างกัน มีรายละเอียดการสร้างแบบสอบ และทดลองใช้ข้อสอบ (Item Tryout) ดังนี้

1. กำหนดองค์ประกอบและพฤติกรรมบ่งชี้ของสมรรถนะทางวิทยาศาสตร์

เครื่องมือที่ใช้ในการวิจัยเป็นแบบสอบเลือกตอบสองระดับด้านสมรรถนะทางวิทยาศาสตร์ รายวิชาวิทยาศาสตร์ชีวภาพ ระดับชั้นมัธยมศึกษาตอนปลาย ซึ่งครอบคลุมเนื้อหา 5 เรื่อง ประกอบด้วยสมรรถนะทางวิทยาศาสตร์ 3 องค์ประกอบ ซึ่งแต่ละองค์ประกอบมี 5 พฤติกรรมบ่งชี้ รายละเอียดดังตาราง 8

ตาราง 8 องค์ประกอบและพฤติกรรมบ่งชี้ของสมรรถนะทางวิทยาศาสตร์

สมรรถนะทางวิทยาศาสตร์	พฤติกรรมบ่งชี้
องค์ประกอบที่ 1 การอธิบายปรากฏการณ์ในเชิง วิทยาศาสตร์	1) ใช้องค์ความรู้ทางวิทยาศาสตร์อธิบายปรากฏการณ์ได้อย่างเหมาะสม
	2) เลือกแบบจำลองทางวิทยาศาสตร์มาใช้เป็นตัวแทนของสิ่งของประกอบคำอธิบายปรากฏการณ์
	3) พยากรณ์การเปลี่ยนแปลงที่อาจเกิดขึ้นได้จริงจากการตีความปรากฏการณ์
	4) เสนอสมมติฐานที่ใช้ในการอธิบายปรากฏการณ์
	5) อธิบายถึงความเป็นไปได้ในการนำองค์ความรู้ทางวิทยาศาสตร์ไปใช้ประโยชน์กับสังคม
องค์ประกอบที่ 2 การประเมินและออกแบบ กระบวนการสืบเสาะหาความรู้ทาง วิทยาศาสตร์	1) ระบุปัญหาที่ต้องการสำรวจในการศึกษาทางวิทยาศาสตร์
	2) บอกได้ว่าประเด็นปัญหาหรือคำถามใดสามารถตรวจสอบได้ด้วยวิธีการทางวิทยาศาสตร์
	3) เสนอวิธีสำรวจตรวจสอบปัญหาทางวิทยาศาสตร์
	4) ประเมินวิธีสำรวจตรวจสอบปัญหาทางวิทยาศาสตร์
	5) ประเมินและอธิบายวิธีการต่าง ๆ ที่นักวิทยาศาสตร์ใช้ยืนยันความน่าเชื่อถือของข้อมูล รวมถึงคำอธิบายปรากฏการณ์ที่ปราศจากความคิดเห็นส่วนตัว

สมรรถนะทางวิทยาศาสตร์	พฤติกรรมบ่งชี้
องค์ประกอบที่ 3	1) แปลงข้อมูลจากรูปแบบการนำเสนอเดิมไปสู่รูปแบบอื่น
การแปลความหมายข้อมูลและการใช้ประจักษ์พยานในเชิงวิทยาศาสตร์	2) วิเคราะห์และตีความหมายข้อมูล และลงข้อสรุปอย่างสมเหตุสมผล
	3) ระบุข้อสันนิษฐาน ประจักษ์พยาน (หลักฐาน) และเหตุผลในเรื่องที่เกี่ยวกับวิทยาศาสตร์
	4) บอกความแตกต่างระหว่างข้อโต้แย้งที่อยู่บนพื้นฐานของประจักษ์พยานทางวิทยาศาสตร์กับข้อโต้แย้งที่ใช้ฐานแนวคิดอื่น ๆ
	5) ประเมินข้อโต้แย้งทางวิทยาศาสตร์และประจักษ์พยานจากแหล่งที่มาที่หลากหลาย เช่น หนังสือพิมพ์ อินเทอร์เน็ต และวารสาร

2. วางแผนการสอบ

การเก็บรวบรวมข้อมูลจะใช้แบบสอบถามสองระดับด้านสมรรถนะทางวิทยาศาสตร์ รายวิชาวิทยาศาสตร์ชีวภาพ ระดับชั้นมัธยมศึกษาตอนปลาย จำนวน 2 ฉบับ ฉบับละ 20 ข้อ ที่มีระดับความมั่นใจที่แตกต่างกัน โดยเก็บข้อมูลนักเรียนระดับชั้นมัธยมศึกษาตอนปลาย จำนวน 750 คน แบ่งเป็น 2 กลุ่ม กลุ่มละ 375 คน นักเรียนกลุ่มที่ 1 จะได้ทำแบบสอบฉบับที่ 1 ซึ่งเป็นแบบสอบที่มีระดับความมั่นใจ 2 ระดับ และนักเรียนกลุ่มที่ 2 จะได้ทำแบบสอบฉบับที่ 2 ซึ่งเป็นแบบสอบที่มีระดับความมั่นใจ 4 ระดับ โดยใช้เวลาในการทำแบบสอบ 60 นาที

3. สร้างแผนผังการสร้างข้อสอบ

จากการศึกษาเนื้อหาวิชาวิทยาศาสตร์ชีวภาพ และองค์ประกอบของสมรรถนะทางวิทยาศาสตร์ ผู้วิจัยกำหนดแผนผังการสร้างข้อสอบ (Table of Specification) เพื่อแสดงรายละเอียดเนื้อหาย่อยในรายวิชาวิทยาศาสตร์ชีวภาพ จำนวน 5 เรื่อง ได้แก่ 1) การลำเลียงสาร 2) การรักษาดุลยภาพของร่างกายมนุษย์ 3) การดำรงชีวิตของพืช 4) พันธุกรรมและวิวัฒนาการ และ 5) ชีวิตในสิ่งแวดล้อม ที่สอดคล้องกับสมรรถนะทางวิทยาศาสตร์ที่ต้องการจะวัดทั้ง 3 องค์ประกอบ ได้แก่ 1) การอธิบายปรากฏการณ์ในเชิงวิทยาศาสตร์ 2) การประเมินและออกแบบกระบวนการสืบเสาะหาความรู้ทางวิทยาศาสตร์ และ 3) การแปลความหมายข้อมูลและการใช้ประจักษ์พยานในเชิงวิทยาศาสตร์ ดังตาราง 9

ตาราง 9 แผนผังการสร้างแบบสอบเลือกตอบด้านสมรรถนะทางวิทยาศาสตร์

เนื้อหา	สมรรถนะทางวิทยาศาสตร์			รวม (ข้อ)	ร้อยละ
	องค์ประกอบที่ 1	องค์ประกอบที่ 2	องค์ประกอบที่ 3		
	การอธิบาย ปรากฏการณ์ใน เชิงวิทยาศาสตร์	การประเมินและ ออกแบบ กระบวนการสืบ เสาะหาความรู้ ทางวิทยาศาสตร์	การแปล ความหมาย ข้อมูลและการใช้ ประจักษ์พยาน ในเชิง วิทยาศาสตร์		
1. การลำเลียงสาร	4 (2)	-	2 (1)	6 (3)	15
2. การรักษาตุลย ภาพของร่างกาย มนุษย์	4 (2)	-	2 (1)	6 (3)	15
3. การดำรงชีวิตของ พืช	-	4 (2)	2 (1)	6 (3)	15
4. พันธุกรรมและ วิวัฒนาการ	4 (2)	2 (1)	4 (2)	10 (5)	25
5. ชีวิตในสิ่งแวดล้อม	4 (2)	2 (1)	6 (3)	12 (6)	30
รวม (ข้อ)	16 (8)	8 (4)	16 (8)	40 (20)	100
ร้อยละ	40	20	40	100	

หมายเหตุ: จำนวนที่สร้าง (จำนวนที่ใช้จริง)

จากตาราง 9 พบว่า องค์ประกอบที่ 1 การอธิบายปรากฏการณ์ในเชิงวิทยาศาสตร์ ประกอบด้วยข้อคำถามจำนวน 8 ข้อ องค์ประกอบที่ 2 การประเมินและออกแบบกระบวนการสืบเสาะหาความรู้ทางวิทยาศาสตร์ ประกอบด้วยข้อคำถามจำนวน 4 ข้อ และองค์ประกอบที่ 3 การแปลความหมายข้อมูลและการใช้ประจักษ์พยานในเชิงวิทยาศาสตร์ ประกอบด้วยข้อคำถามจำนวน 8 ข้อ ซึ่งคิดเป็นร้อยละ 40, 20 และ 40 ตามลำดับ ดังนั้นแบบสอบเลือกตอบด้านสมรรถนะทาง

วิทยาศาสตร์ รายวิชาวิทยาศาสตร์ชีวภาพ ระดับชั้นมัธยมศึกษาตอนปลาย ประกอบด้วยข้อคำถาม จำนวน 20 ข้อ

4. สร้างแบบสอบ

ผู้วิจัยต้องการสร้างแบบสอบเลือกตอบด้านสมรรถนะทางวิทยาศาสตร์ รายวิชาวิทยาศาสตร์ชีวภาพ ระดับชั้นมัธยมศึกษาตอนปลาย จำนวน 20 ข้อ ซึ่งเป็นแบบสอบเลือกตอบ 4 ตัวเลือก แต่เพื่อมีการคัดเลือกข้อสอบที่ไม่มีคุณภาพออก ผู้วิจัยจึงต้องสร้างข้อสอบเป็น 2 เท่า นั่นคือจำนวน 40 ข้อ โดยผู้วิจัยออกข้อสอบตามแผนผังการสร้างข้อสอบ จากนั้นทบทวนข้อสอบด้านความชัดเจนและภาษาที่ใช้ในการออกข้อสอบอีกครั้ง

5. ตรวจสอบคุณภาพของแบบสอบ

ในการตรวจสอบคุณภาพของแบบสอบเลือกตอบด้านสมรรถนะทางวิทยาศาสตร์ โดยอาศัยผู้เชี่ยวชาญเพื่อตรวจสอบความตรงเชิงเนื้อหา (Content Validity) โดยผู้เชี่ยวชาญทั้ง 3 ท่าน ได้แก่

- | | |
|---|--------------|
| 1) ผู้เชี่ยวชาญด้านการวิจัย | จำนวน 1 ท่าน |
| 2) ผู้เชี่ยวชาญด้านการวัดและประเมินผล | จำนวน 1 ท่าน |
| 3) ผู้เชี่ยวชาญด้านการสอนรายวิชาวิทยาศาสตร์ | จำนวน 1 ท่าน |

การตรวจสอบความตรงเชิงเนื้อหา ผู้เชี่ยวชาญจะต้องทำเครื่องหมาย P ลงในช่องระดับการพิจารณา ดังนี้

- | | | |
|----|------------------|---|
| +1 | เมื่อแน่ใจว่า | ข้อคำถามนั้นสอดคล้องกับสมรรถนะทางวิทยาศาสตร์ |
| 0 | เมื่อไม่แน่ใจว่า | ข้อคำถามนั้นสอดคล้องกับสมรรถนะทางวิทยาศาสตร์ |
| -1 | เมื่อแน่ใจว่า | ข้อคำถามนั้นไม่สอดคล้องกับสมรรถนะทางวิทยาศาสตร์ |

เมื่อผู้เชี่ยวชาญทั้ง 3 ท่านพิจารณาความสอดคล้องระหว่างข้อคำถามกับวัตถุประสงค์ ซึ่งในงานวิจัยนี้คือพฤติกรรมบ่งชี้ของสมรรถนะทางวิทยาศาสตร์แล้ว จากนั้นจะพิจารณาความตรงเชิงเนื้อหาโดยคำนวณจากสูตร IOC ดังนี้

$$IOC = \frac{\sum R}{N}$$

- | | | | |
|-------|----------|-----|---------------------------------|
| เมื่อ | $\sum R$ | คือ | ผลรวมความคิดเห็นของผู้เชี่ยวชาญ |
| | N | คือ | จำนวนของผู้เชี่ยวชาญ |

เกณฑ์การแปลความหมายค่า IOC (ฉันทุฎฐกรณั หลาวทอง, 2559)

$IOC \geq 0.50$ หมายความว่า ข้อสอบข้อนั้นวัดได้สอดคล้องกับจุดประสงค์

$IOC < 0.50$ หมายความว่า ข้อสอบข้อนั้นวัดไม่สอดคล้องกับจุดประสงค์

6. แก้ไขข้อสอบตามคำแนะนำของผู้เชี่ยวชาญ ตัดข้อสอบที่ไม่มีคุณภาพ และดำเนินการพิมพ์ข้อสอบ

หลังจากตรวจสอบคุณภาพของแบบสอบด้านความตรงเชิงเนื้อหาโดยผู้เชี่ยวชาญทั้ง 3 ท่าน และแก้ไขตามคำแนะนำของผู้เชี่ยวชาญแล้ว ผู้วิจัยจะคัดเลือกข้อสอบที่มีค่า IOC มากกว่า 0.50 ขึ้นไป จำนวน 40 ข้อ และดำเนินการพิมพ์แบบสอบฉบับจริงเพื่อนำไปดำเนินการใช้แบบสอบ (Tryout) วิเคราะห์คุณภาพของแบบสอบ และนำแบบสอบไปใช้ในงานวิจัยต่อไป

7. ดำเนินการทดลองใช้แบบสอบ (Tryout)

การดำเนินการทดลองใช้แบบสอบ เพื่อวิเคราะห์คุณภาพของแบบสอบ โดยนำแบบสอบเลือกตอบด้านสมรรถนะทางวิทยาศาสตร์ รายวิชาวิทยาศาสตร์ชีวภาพ ระดับชั้นมัธยมศึกษาตอนปลาย จำนวน 40 ข้อ ไปทดลองใช้กับนักเรียน โรงเรียนประชาสงเคราะห์วิทยา จำนวน 40 คน โดยใช้เวลาในการทำแบบสอบ 60 นาที และควบคุมให้นักเรียนเริ่มทำและส่งแบบสอบเวลาเดียวกัน

8. วิเคราะห์คุณภาพของแบบสอบ

การวิเคราะห์คุณภาพของแบบสอบเลือกตอบด้านสมรรถนะทางวิทยาศาสตร์ รายวิชาวิทยาศาสตร์ชีวภาพ ระดับชั้นมัธยมศึกษาตอนปลาย ตามทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory) โดยวิเคราะห์ความยากและอำนาจจำแนก คะแนนเฉลี่ยรายข้อและทั้งฉบับ ความแปรปรวนของคะแนนรายข้อและทั้งฉบับ และความเที่ยง หลักจากการวิเคราะห์พบว่า มีข้อสอบที่ไม่ผ่านเกณฑ์จำนวน 17 ข้อ ได้แก่ ข้อ 1, ข้อ 4, ข้อ 5, ข้อ 8, ข้อ 9, ข้อ 10, ข้อ 12, ข้อ 14, ข้อ 17, ข้อ 19, ข้อ 20, ข้อ 22, ข้อ 26, ข้อ 32, ข้อ 33, ข้อ 35 และข้อ 40 ดังนั้น จึงตัดข้อสอบดังกล่าวออก

ผลการวิเคราะห์ค่าความยากและอำนาจจำแนกของแบบสอบเลือกตอบด้านสมรรถนะทางวิทยาศาสตร์ แสดงดังตาราง 10

ตาราง 10 ผลการวิเคราะห์ค่าความยากและอำนาจจำแนกของแบบสอบเลือกตอบด้านสมรรถนะทางวิทยาศาสตร์

ข้อที่	ค่าความยาก	ค่าอำนาจจำแนก	sig	ข้อที่	ค่าความยาก	ค่าอำนาจจำแนก	sig
1	0.20	-0.04	1.00	21	0.55	0.46*	0.00
2	0.65	0.40*	0.01	22	0.05	-0.22	1.00
3	0.75	0.33*	0.04	23	0.58	0.39*	0.02
4	0.28	-0.08	1.00	24	0.45	0.35*	0.01
5	0.45	0.12	0.45	25	0.45	0.39*	0.01
6	0.53	0.48*	0.00	26	0.13	0.01	0.97
7	0.60	0.35*	0.03	27	0.60	0.33*	0.04
8	0.55	0.15	0.35	28	0.50	0.36*	0.02
9	0.08	-0.03	1.00	29	0.60	0.40*	0.01
10	0.15	0.03	0.86	30	0.45	0.43*	0.00
11	0.38	0.48*	0.00	31	0.65	0.34*	0.03
12	0.78	-0.15	1.00	32	0.23	-0.18	1.00
13	0.35	0.38*	0.02	33	0.18	0.10	0.55
14	0.40	0.10	0.55	34	0.50	0.49*	0.00
15	0.53	0.46*	0.00	35	0.08	0.02	0.92
16	0.48	0.62*	0.00	36	0.60	0.32*	0.04
17	0.20	-0.23	1.00	37	0.55	0.36*	0.02
18	0.65	0.46*	0.01	38	0.45	0.38*	0.02
19	0.13	-0.25	1.00	39	0.58	0.42*	0.01
20	0.13	-0.23	1.00	40	0.05	0.09	0.59

หมายเหตุ: ข้อที่แสดงด้วยสี หมายความว่า ใช้ไม่ได้/ตัดทิ้ง

9. คัดเลือกข้อสอบ

ในการคัดเลือกข้อสอบ โดยคัดเลือกตามองค์ประกอบของสมรรถนะทางวิทยาศาสตร์ทั้ง 3 องค์ประกอบ และครอบคลุมเนื้อหารายวิชาวิทยาศาสตร์ชีวภาพทั้ง 5 เรื่อง จำนวน 20 ข้อ เพื่อนำไปใช้ในการเก็บข้อมูลจริง

การเก็บรวบรวมข้อมูล

การเก็บรวบรวมข้อมูลผลการตอบและการตรวจให้คะแนนความรู้บางส่วน มีขั้นตอนในการเก็บรวบรวมข้อมูล ดังนี้

1. ส่งหนังสือไปยังผู้อำนวยการโรงเรียนต่าง ๆ เพื่อขอเก็บข้อมูล และประสานงานกับครูผู้สอนรายวิชาวิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนปลาย เพื่อแนะแนวทางในการทำแบบสอบ

2. นำแบบสอบเลือกตอบสองระดับด้านสมรรถนะทางวิทยาศาสตร์ ที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่แตกต่างกัน จำนวน 20 ข้อ มาทดสอบกับนักเรียนจำนวน 750 คน ให้ความเวลาในการทำแบบสอบ 60 นาที ซึ่งเป็นการทดสอบผ่านระบบออนไลน์ด้วย Google Form โดยผู้วิจัยมีการชี้แจงคำสั่ง วิธีการตอบ และการตรวจให้คะแนนความรู้บางส่วนอย่างละเอียด กลุ่มตัวอย่างวิจัยแบ่งเป็น 2 กลุ่ม ดังนี้

กลุ่มที่ 1 ผู้สอบจะต้องตอบแบบสอบเลือกตอบสองระดับด้านสมรรถนะทางวิทยาศาสตร์ ที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 2 ระดับ ได้แก่ มั่นใจ และไม่มั่นใจ

กลุ่มที่ 2 ผู้สอบจะต้องตอบแบบสอบเลือกตอบสองระดับด้านสมรรถนะทางวิทยาศาสตร์ ที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 4 ระดับ ได้แก่ มั่นใจมาก มั่นใจ ไม่มั่นใจ และ เด่า

3. นำผลการตอบมาตรวจให้คะแนนความรู้บางส่วน ซึ่งวิธีการตอบและตรวจให้คะแนนความรู้บางส่วนด้วยวิธีการตอบโดยบอกระดับความมั่นใจ ประกอบด้วย 2 วิธี ดังนี้

3.1 วิธีการตอบและตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 2 ระดับ โดยในการตอบวิธีนี้ผู้สอบจะตอบลงในกระดาษคำตอบที่มี 2 ส่วนคือ คำตอบระดับเนื้อหา (Content Tier) และคำตอบระดับความมั่นใจ (Confidence Tier) ที่ประกอบด้วยความมั่นใจ 2 ระดับ ได้แก่ มั่นใจ และไม่มั่นใจ วิธีการตอบและตรวจให้คะแนนแสดงดังตัวอย่าง

ข้อสอบข้อที่ 0 คำตอบที่ถูกต้องคือ ข้อ ข ผู้สอบตอบข้อ ข และระบุความมั่นใจ “มั่นใจมาก” จะได้คะแนน 2 คะแนน

ข้อ	ตัวเลือก				ระดับความมั่นใจ		คะแนน
	ก	ข	ค	ง	มั่นใจ	ไม่มั่นใจ	
0		ข			มั่นใจ		2

ข้อสอบข้อที่ 0 คำตอบที่ถูกต้องคือ ข้อ ข ผู้สอบตอบข้อ ข และระดับความมั่นใจ “ไม่มั่นใจ”
จะได้คะแนน 1 คะแนน

ข้อ	ตัวเลือก				ระดับความมั่นใจ		คะแนน
	ก	ข	ค	ง	มั่นใจ	ไม่มั่นใจ	
0		ı				ı	1

ข้อสอบข้อที่ 0 คำตอบที่ถูกต้องคือ ข้อ ข ถ้าผู้สอบตอบผิดคือ ไม่ได้เลือกข้อ ข จะได้ 0
คะแนน ไม่ว่าจะระบุความมั่นใจในการตอบระดับใด

ข้อ	ตัวเลือก				ระดับความมั่นใจ		คะแนน
	ก	ข	ค	ง	มั่นใจ	ไม่มั่นใจ	
0	ı				ı		0

3.2 วิธีการตอบและตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ
4 ระดับ โดยในการตอบวิธีนี้ผู้สอบจะตอบลงในกระดาษคำตอบที่มี 2 ส่วนคือ คำตอบระดับเนื้อหา
(Content Tier) และคำตอบระดับความมั่นใจ (Confidence Tier) ที่ประกอบด้วยความมั่นใจ 4
ระดับ ได้แก่ มั่นใจมาก มั่นใจ ไม่มั่นใจ และเดา วิธีการตอบและตรวจให้คะแนนแสดงดังตัวอย่าง

ข้อสอบข้อที่ 0 คำตอบที่ถูกต้องคือ ข้อ ข ผู้สอบตอบข้อ ข และระดับความมั่นใจ “มั่นใจมาก”
จะได้คะแนน 4 คะแนน

ข้อ	ตัวเลือก				ระดับความมั่นใจ				คะแนน
	ก	ข	ค	ง	มั่นใจมาก	มั่นใจ	ไม่มั่นใจ	เดา	
0		ı			ı				4

ข้อสอบข้อที่ 0 คำตอบที่ถูกต้องคือ ข้อ ข ผู้สอบตอบข้อ ข และระดับความมั่นใจ “มั่นใจ” จะ
ได้คะแนน 3 คะแนน

ข้อ	ตัวเลือก				ระดับความมั่นใจ				คะแนน
	ก	ข	ค	ง	มั่นใจมาก	มั่นใจ	ไม่มั่นใจ	เดา	
0		ı				ı			3

ข้อสอบข้อที่ 0 คำตอบที่ถูกต้องคือ ข้อ ข ผู้สอบตอบข้อ ข และระบุความมั่นใจ “ไม่มั่นใจ”
จะได้คะแนน 2 คะแนน

ข้อ	ตัวเลือก				ระดับความมั่นใจ				คะแนน
	ก	ข	ค	ง	มั่นใจมาก	มั่นใจ	ไม่มั่นใจ	เดา	
0		ı					ı		2

ข้อสอบข้อที่ 0 คำตอบที่ถูกต้องคือ ข้อ ข ผู้สอบตอบข้อ ข และระบุความมั่นใจ “เดา” จะได้
คะแนน 1 คะแนน

ข้อ	ตัวเลือก				ระดับความมั่นใจ				คะแนน
	ก	ข	ค	ง	มั่นใจมาก	มั่นใจ	ไม่มั่นใจ	เดา	
0		ı						ı	1

ข้อสอบข้อที่ 0 คำตอบที่ถูกต้องคือ ข้อ ข ถ้าผู้สอบตอบผิดคือ ไม่ได้เลือกข้อ ข จะได้ 0
คะแนน ไม่ว่าจะระบุความมั่นใจในการตอบระดับใด

ข้อ	ตัวเลือก				ระดับความมั่นใจ				คะแนน
	ก	ข	ค	ง	มั่นใจมาก	มั่นใจ	ไม่มั่นใจ	เดา	
0	ı				ı				0

การวิเคราะห์ข้อมูล

การวิเคราะห์ข้อมูลงานวิจัยนี้ เพื่อเปรียบเทียบความยากและอำนาจจำแนกของแบบสอบ
ความตรงเชิงโครงสร้างของแบบสอบ ความเที่ยงของแบบสอบ ฟังก์ชันสารสนเทศของแบบสอบ และ
ความแม่นยำในการประมาณค่าความสามารถของผู้สอบที่ได้จากการตอบแบบสอบเลือกตอบสอง
ระดับด้านสมรรถนะทางวิทยาศาสตร์ที่มีระดับความมั่นใจที่แตกต่างกันคือ ความมั่นใจ 2 ระดับ และ
ความมั่นใจ 4 ระดับ ซึ่งมีรายละเอียดในการวิเคราะห์ข้อมูล ดังนี้

1. การวิเคราะห์ความยากและอำนาจจำแนกของแบบสอบ

การวิเคราะห์ความยากและอำนาจจำแนกของแบบสอบที่ได้จากวิธีการตรวจให้คะแนน
ความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่แตกต่างกันคือ ความมั่นใจ 2 ระดับ และความมั่นใจ
4 ระดับ โดยวิเคราะห์ด้วยโปรแกรมสำเร็จรูป IRTPRO4 Student และวิเคราะห์ด้วย Generalized
Partial Credit Model (G-PCM)

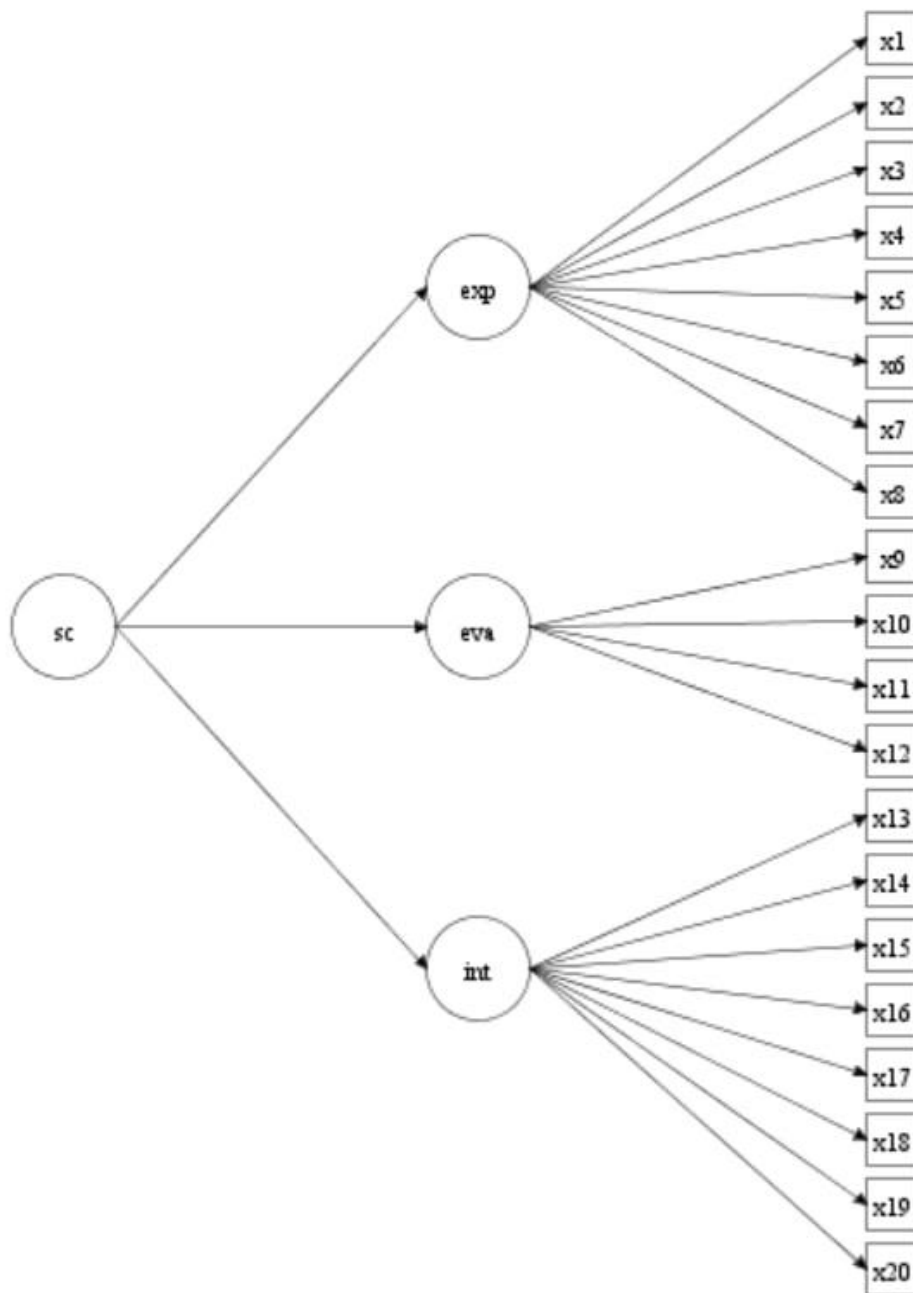
ในการนำทฤษฎีการตอบสนองข้อสอบไปใช้ จำเป็นต้องพิจารณาถึงข้อตกลงเบื้องต้นก่อน ด้วยการตรวจสอบความเป็นเอกมิติ (Unidimensionality) และความเป็นอิสระ (Local Independence) อย่างไรก็ตาม Warm (1978) กล่าวว่า ถ้าแบบสอบมีความเป็นเอกมิติแล้ว แสดงว่าแบบสอบนั้นมีความเป็นอิสระต่อกัน แต่ถ้าแบบสอบมีความเป็นอิสระต่อกันแล้วก็ยังไม่เพียงพอที่จะสรุปได้ว่าแบบสอบนั้นมีความเป็นเอกมิติ ซึ่งสอดคล้องกับ Dorans (1985), Hambleton (1989) และ Hambleton et al. (1991) ที่ได้กล่าวว่า ข้อตกลงเบื้องต้นหลักของทฤษฎีการตอบสนองข้อสอบคือ ความเป็นเอกมิติ หากข้อสอบมีความเป็นเอกมิติแล้วก็แสดงว่าข้อตกลงเบื้องต้นอื่น ๆ จะได้รับการยอมรับด้วย เช่น ความเป็นอิสระ เป็นต้น

ดังนั้นในการวิจัยครั้งนี้ ผู้วิจัยจึงทำการตรวจสอบเฉพาะความเป็นเอกมิติเท่านั้น โดยใช้วิธีการวิเคราะห์หองค์ประกอบเชิงสำรวจ (Exploratory Factor Analysis: EFA) ด้วยวิธีวิเคราะห์หองค์ประกอบหลัก (Principle Component Analysis) โดยใช้โปรแกรมสำเร็จรูป SPSS เพื่อตรวจสอบว่า แบบสอบเลือกตอบด้านสมรรถนะทางวิทยาศาสตร์ รายวิชาวิทยาศาสตร์ชีวภาพ ระดับชั้นมัธยมศึกษาตอนปลายที่สร้างขึ้นมีเพียง 1 หองค์ประกอบเท่านั้น ซึ่งมีเกณฑ์พิจารณาความเป็นเอกมิติ 2 เกณฑ์คือ 1) การพิจารณาค่าความแปรปรวนของหองค์ประกอบแรก ซึ่งถ้าสามารถอธิบายความแปรปรวนทั้งหมดได้อย่างน้อยร้อยละ 20 จะบ่งบอกถึงความเป็นเอกมิติ (Reckase, 1979) และ 2) การพิจารณาค่าไอเกน (Eigenvalue: λ) คือ หากค่าไอเกนตัวแรกมีค่ามากกว่าตัวที่สองและตัวอื่น ๆ อย่างเห็นได้ชัด หรือพิจารณาจากการคำนวณอัตราส่วนของค่าไอเกน (Eigen Ratio: ER) ของหองค์ประกอบที่ 1 กับหองค์ประกอบที่ 2 (λ_1/λ_2) ที่ได้จากการวิเคราะห์ โดยค่าที่ได้ไม่ควรต่ำกว่า 3 (Lord & Novick, 1968) โดยทั้งสองเกณฑ์จะแสดงว่าแบบสอบนั้นมีการวัดเพียงคุณลักษณะเดียว (Single Dominant Factor) ซึ่งบ่งบอกถึงความเป็นเอกมิติ

2. การวิเคราะห์ความตรงเชิงโครงสร้างของแบบสอบ

ในงานวิจัยนี้เลือกใช้การวิเคราะห์หองค์ประกอบเชิงยืนยัน (Confirmatory Factor Analysis) ในการพิจารณาความสอดคล้องระหว่างแบบสอบกับคุณลักษณะที่ต้องการวัด โดยการประมาณค่าพารามิเตอร์ของโมเดลใช้โปรแกรมสำเร็จรูป Mplus และเป็นการประมาณค่าความน่าจะเป็นสูงสุด (Maximum Likelihood) การตรวจสอบความสอดคล้องระหว่างโมเดลและข้อมูล ทำได้โดยการพิจารณาค่าไคสแควร์ (χ^2), ไคสแควร์ต่อองศาอิสระ (χ^2/df), RMSEA, CFI และ SRMR โดยเกณฑ์ที่ใช้ในการพิจารณาความสอดคล้องระหว่างโมเดลและข้อมูลคือ χ^2 ต้องมีค่า p-value มากกว่า 0.05 χ^2/df ต้องไม่เกิน 2, RMSEA และ SRMR ต้องไม่เกิน 0.05 และ CFI ต้องมากกว่า 0.97 ดังตาราง 10 จากนั้นเปรียบเทียบความสอดคล้องระหว่างโมเดลและข้อมูล โดยการพิจารณาค่า χ^2 , χ^2/df , RMSEA, CFI และ SRMR ของแบบสอบเลือกตอบสองระดับด้านสมรรถนะทาง

วิทยาศาสตร์ที่มีระดับความมั่นใจที่แตกต่างกันคือ ความมั่นใจ 2 ระดับ และความมั่นใจ 4 ระดับ ซึ่งโครงสร้างของแบบสอบสำหรับการวิเคราะห์องค์ประกอบเชิงยืนยันของสมรรถนะทางวิทยาศาสตร์ทั้ง 3 องค์ประกอบ ดังนั้น คุณลักษณะที่ต้องการวัด เท่ากับ 3 และตัวแปรสังเกตได้ เป็นข้อสอบจำนวน 20 ข้อ ($X_1 - X_{20}$) และโครงสร้างของโมเดล แสดงดังภาพ 7



ภาพ 7 โครงสร้างของแบบสอบสำหรับการวิเคราะห์องค์ประกอบเชิงยืนยัน

ตาราง 11 เกณฑ์การพิจารณาดัชนีความสอดคล้องระหว่างโมเดลและข้อมูล

ดัชนีความสอดคล้อง	ค่าที่ยอมรับว่าสอดคล้อง
χ^2	$0.05 < p - value < 1$
χ^2/df	$0 < \chi^2/df \leq 2$
RMSEA	$0 \leq RMSEA \leq 0.05$
SRMR	$0 \leq SRMR \leq 0.05$
CFI	$0.97 \leq CFI \leq 1$

ที่มา: สุวิมล ตีรกาพันธ์ (2555)

3. การวิเคราะห์ความเที่ยงของแบบสอบ

การวิเคราะห์ความเที่ยงของแบบสอบได้จากการวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่ต่างกันคือ ความมั่นใจ 2 ระดับ และความมั่นใจ 4 ระดับ โดยใช้ความเที่ยงแบบความสอดคล้องภายใน วิธีสัมประสิทธิ์แอลฟาครอนบาค (Cronbach's Alpha) คำนวณได้จากสูตรดังต่อไปนี้

$$\alpha = \left[\frac{k}{k-1} \right] \left[1 - \frac{\sum S_i^2}{S_x^2} \right]$$

เมื่อ α	คือ	สัมประสิทธิ์ความเที่ยงของแบบสอบ
S_i^2	คือ	ความแปรปรวนของคะแนนข้อที่
S_x^2	คือ	ความแปรปรวนของคะแนนรวม
k	คือ	จำนวนข้อสอบ

4. การวิเคราะห์ฟังก์ชันสารสนเทศของแบบสอบ

การวิเคราะห์ฟังก์ชันสารสนเทศของแบบสอบที่ได้จากการวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่ต่างกันคือ ความมั่นใจ 2 ระดับ และความมั่นใจ 4 ระดับ โดยการวิเคราะห์ด้วยโปรแกรมสำเร็จรูป IRTPRO4 Student เพื่ออธิบายความสัมพันธ์ระหว่างความสามารถของผู้สอบและโอกาสในการเลือกตอบแต่ละรายการที่กำหนด ซึ่งการวิเคราะห์ฟังก์ชันสารสนเทศของแบบสอบจะวิเคราะห์ด้วย Generalized Partial Credit Model (G-PCM) สมการของ G-PCM (Muraki, 1992) มีดังนี้

$$C_{jk}(\theta) = P_{jk|k-1,k}(\theta) = \frac{P_{jk}(\theta)}{P_{j,k-1}(\theta) + P_{jk}(\theta)} = \frac{\exp[a_j(\theta - \delta_{jk})]}{1 + \exp[a_j(\theta - \delta_{jk})]}$$

$$P_{jk}(\theta) = \frac{C_{jk}}{1 - C_{jk}} P_{j,k-1,k}(\theta)$$

เมื่อ	k	คือ	ลำดับชั้นคะแนนที่ 1, 2, 3, ..., m_j
	$P_{jk}(\theta)$	คือ	ความน่าจะเป็นของคนที่มีความสามารถจะตอบได้คะแนน k
	$\frac{C_{jk}}{1 - C_{jk}}$	คือ	อัตราส่วนความน่าจะเป็นของสองเงื่อนไข ซึ่งอาจจะแสดงได้ในรูป $\exp[a_j(\theta - \delta_{jk})]$
	a_j	คือ	ค่าพารามิเตอร์อำนาจจำแนกของข้อ j
	b_{jk}	คือ	ค่าพารามิเตอร์ความยากของชั้นการตอบที่ k ในข้อ j เมื่อ $k = 1, 2, 3, \dots, m_j$

นอกจากนี้ G-PCM สามารถวิเคราะห์ฟังก์ชันสารสนเทศของแบบสอบ (Test Information) ซึ่งเป็นดัชนีที่บ่งชี้คุณภาพของแบบสอบทั้งฉบับ ดังสมการ

$$I(\theta) = \sum_{i=1}^k I_i(\theta)$$

เมื่อ	$I(\theta)$	คือ	ค่าฟังก์ชันสารสนเทศของแบบสอบ หรือค่าสารสนเทศที่ได้รับจากแบบสอบสำหรับผู้ที่มีความสามารถ θ
-------	-------------	-----	---

บทที่ 4

ผลการวิจัย

การวิจัยเรื่อง การเปรียบเทียบคุณภาพของแบบสอบเลือกตอบสองระดับด้านสมรรถนะทางวิทยาศาสตร์ที่มีระดับความมั่นใจที่แตกต่างกัน โดยประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบแบบตรวจให้คะแนนความรู้บางส่วนแบบทั่วไป มีวัตถุประสงค์ 4 ประการ ได้แก่ 1) เพื่อเปรียบเทียบความยากและอำนาจจำแนกของแบบสอบ 2) เพื่อเปรียบเทียบความตรงเชิงโครงสร้างของแบบสอบ 3) เพื่อเปรียบเทียบความเที่ยงของแบบสอบ และ 4) เพื่อเปรียบเทียบฟังก์ชันสารสนเทศของแบบสอบ โดยผู้วิจัยเสนอผลการวิเคราะห์ข้อมูลออกเป็น 5 ตอน ดังนี้

ตอนที่ 1 ผลการวิเคราะห์ความยากและอำนาจจำแนกของแบบสอบ

ตอนที่ 2 ผลการวิเคราะห์ความตรงเชิงโครงสร้างของแบบสอบ

ตอนที่ 3 ผลการวิเคราะห์ความเที่ยงของแบบสอบ

ตอนที่ 4 ผลการวิเคราะห์ฟังก์ชันสารสนเทศของแบบสอบ

ตอนที่ 5 ผลการเปรียบเทียบคุณภาพของแบบสอบ

ตอนที่ 1 ผลการวิเคราะห์ความยากและอำนาจจำแนกของแบบสอบ

1.1 ผลการตรวจสอบข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ

การตรวจสอบข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบครั้งนี้ ผู้วิจัยทำการตรวจสอบเฉพาะความเป็นเอกมิติ (Unidimensionality) ใช้วิธีการวิเคราะห์องค์ประกอบเชิงสำรวจ (Exploratory Factor Analysis: EFA) ด้วยวิธีวิเคราะห์องค์ประกอบหลัก (Principle Component Analysis) โดยใช้โปรแกรมสำเร็จรูป SPSS ของแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่แตกต่างกัน พบว่า แบบสอบทั้ง 2 ฉบับ ผ่านเกณฑ์ข้อตกลงเบื้องต้นด้านความเป็นเอกมิติของการวัด เมื่อใช้เกณฑ์ค่าความแปรปรวน ซึ่งสามารถอธิบายความแปรปรวนทั้งหมดได้อย่างน้อยร้อยละ 20 จะบ่งบอกถึงความเป็นเอกมิติ พบว่า แบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 2 ระดับและ 4 ระดับมีค่าร้อยละความแปรปรวนเท่ากับ 27.06 และ 21.39 ตามลำดับ

ในกรณีใช้เกณฑ์อัตราส่วนของค่าไอเกนขององค์ประกอบที่ 1 กับองค์ประกอบที่ 2 (λ_1/λ_2) ที่ได้จากวิธีการวิเคราะห์ โดยค่าที่ได้ไม่ควรต่ำกว่า 3 จะบ่งบอกถึงความเป็นเอกมิติ พบว่า แบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 2 ระดับ มีค่าอัตราส่วนของ

ค่าไอเกนขององค์ประกอบที่ 1 กับองค์ประกอบที่ 2 เท่ากับ 3.13 ซึ่งเป็นไปตามเกณฑ์ ส่วนแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 4 ระดับ มีค่าอัตราส่วนของค่าไอเกนขององค์ประกอบที่ 1 กับองค์ประกอบที่ 2 เท่ากับ 2.88 ซึ่งไม่เป็นไปตามเกณฑ์ โดยผลการวิเคราะห์ข้อมูล แสดงดังตาราง 12

ตาราง 12 ผลการตรวจสอบข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ

วิธีการตรวจให้คะแนน	ร้อยละความแปรปรวน	อัตราส่วนของค่าไอเกน
ระดับความมั่นใจ 2 ระดับ	27.06	3.13
ระดับความมั่นใจ 4 ระดับ	21.39	2.88

เนื่องจากแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 2 ระดับและ 4 ระดับ ผ่านเกณฑ์การตรวจสอบข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบครั้งนี้อย่างน้อย 1 เกณฑ์ ดังนั้นจึงสามารถนำแบบสอบทั้ง 2 ฉบับ ไปวิเคราะห์สารสนเทศตามทฤษฎีการตอบสนองข้อสอบได้

1.2 ผลการวิเคราะห์ความยากและอำนาจจำแนกของแบบสอบ

การวิเคราะห์ความยากและอำนาจจำแนกของแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่แตกต่างกัน ด้วยโปรแกรมสำเร็จรูป IRTPRO4 Student และวิเคราะห์ด้วย Generalized Partial Credit Model (G-PCM) โดยมีเกณฑ์การพิจารณาคือ ค่าความยาก (b) ควรอยู่ระหว่าง -4.0 ถึง +4.0 และค่าอำนาจจำแนก (a) ควรอยู่ระหว่าง 0 ถึง ∞ หมายความว่า ข้อสอบสามารถนำไปใช้ได้ (ณภัทร ชัยมงคล, 2559) พบว่า แบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 2 ระดับ มีค่าความยากระหว่าง -0.35 ถึง 8.75 ค่าอำนาจจำแนกระหว่าง 0.06 ถึง 1.19 และมีค่าความยากเฉลี่ยและค่าอำนาจจำแนกเฉลี่ย เท่ากับ 1.32 และ 0.69 ตามลำดับ แสดงว่า ข้อสอบค่อนข้างยาก และอำนาจจำแนกค่อนข้างต่ำ โดยมีข้อที่ไม่ผ่านเกณฑ์คือ ข้อ 20 ส่วนแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 4 ระดับ มีค่าความยากระหว่าง 1.38 ถึง 5.05 ค่าอำนาจจำแนกระหว่าง 0.12 ถึง 0.61 และมีค่าความยากเฉลี่ยและค่าอำนาจจำแนกเฉลี่ย เท่ากับ 2.72 และ 0.26 ตามลำดับ แสดงว่า ข้อสอบค่อนข้างยาก และอำนาจจำแนกค่อนข้างต่ำ โดยมีข้อที่ไม่ผ่านเกณฑ์คือ ข้อ 12 ผลการวิเคราะห์ข้อมูล แสดงดังตาราง 13 และ 14

ตาราง 13 ผลการวิเคราะห์ความยากและอำนาจจำแนกของแบบสอบที่ได้จากวิธีการตรวจให้คะแนน
ความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่แตกต่างกัน

ข้อ	ระดับความมั่นใจ 2 ระดับ		ระดับความมั่นใจ 4 ระดับ	
	<i>b</i>	<i>a</i>	<i>b</i>	<i>a</i>
1	0.77	0.71	1.92	0.28
2	0.67	0.94	2.58	0.21
3	0.64	0.83	2.52	0.21
4	0.31	0.71	3.18	0.23
5	1.44	0.50	1.82	0.33
6	0.82	1.02	3.70	0.18
7	-0.35	0.65	2.40	0.20
8	1.24	0.85	2.28	0.21
9	0.67	0.89	2.60	0.30
10	0.24	0.65	2.76	0.24
11	0.51	1.19	2.41	0.29
12	3.72	0.24	5.05	0.12
13	0.46	0.91	3.76	0.22
14	0.20	0.67	3.38	0.18
15	1.24	0.75	3.54	0.17
16	0.62	0.69	1.55	0.30
17	0.33	0.84	1.38	0.61
18	3.11	0.14	3.39	0.21
19	1.07	0.65	1.56	0.48
20	8.75	0.06	2.59	0.30

หมายเหตุ: *b* คือ ความยาก และ *a* คือ อำนาจจำแนก

ตาราง 14 ผลการวิเคราะห์ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของความยากและอำนาจจำแนกของแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่แตกต่างกัน

วิธีการตรวจให้คะแนน	n	ความยาก		อำนาจจำแนก	
		M	SD	M	SD
ระดับความมั่นใจ 2 ระดับ	375	1.32	1.99	0.69	0.28
ระดับความมั่นใจ 4 ระดับ	375	2.72	0.91	0.26	0.11

จากตาราง 14 แสดงว่า แบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 2 ระดับ มีความยากของข้อสอบมากกว่าและสามารถจำแนกผู้สอบได้ดีกว่าแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 4 ระดับ

1.3 ผลการวิเคราะห์ความแตกต่างระหว่างค่าความยากและอำนาจจำแนกเฉลี่ยของแบบสอบ

การวิเคราะห์ความแตกต่างระหว่างค่าความยากและอำนาจจำแนกเฉลี่ยของแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่แตกต่างกัน สามารถวิเคราะห์ได้โดยการวิเคราะห์ความแตกต่างของค่าเฉลี่ย (Independent T-test) ด้วยโปรแกรมสำเร็จรูป SPSS

1.3.1 ผลการวิเคราะห์ความแตกต่างระหว่างค่าความยากเฉลี่ยของแบบสอบ

จากการวิเคราะห์ความแตกต่างระหว่างค่าความยากเฉลี่ยของแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่แตกต่างกัน โดยใช้การวิเคราะห์ความแตกต่างของค่าเฉลี่ย (Independent T-test) พบว่า วิธีการให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่แตกต่างกันมีค่าความยากของแบบสอบแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ผลการวิเคราะห์ข้อมูล แสดงดังตาราง 15

ตาราง 15 ผลการวิเคราะห์ความแตกต่างของค่าเฉลี่ยของค่าความยากเฉลี่ยของแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่แตกต่างกัน

แหล่งของความแปรปรวน	Levene's Test		T-test				
	F	sig	t	df	sig	M	SD
ค่าความยาก	1.58	0.22	-2.85	38	.01	-1.40	0.49

1.3.2 ผลการวิเคราะห์ความแตกต่างระหว่างค่าอำนาจจำแนกเฉลี่ยของแบบสอบ

จากการวิเคราะห์ความแตกต่างระหว่างค่าอำนาจจำแนกเฉลี่ยของแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่แตกต่างกัน โดยใช้การวิเคราะห์ความแตกต่างของค่าเฉลี่ย (Independent T-test) พบว่า วิธีการให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่แตกต่างกันมีค่าอำนาจจำแนกของแบบสอบแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ผลการวิเคราะห์ข้อมูล แสดงดังตาราง 16

ตาราง 16 ผลการวิเคราะห์ความแตกต่างของค่าเฉลี่ยของค่าอำนาจจำแนกเฉลี่ยของแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่แตกต่างกัน

แหล่งของความแปรปรวน	Levene's Test				T-test		
	F	sig	t	df	sig	M	SD
ค่าอำนาจจำแนก	6.71	0.01	6.33	24.81	.00	0.43	0.07

ตอนที่ 2 ผลการวิเคราะห์ความตรงเชิงโครงสร้างของแบบสอบ

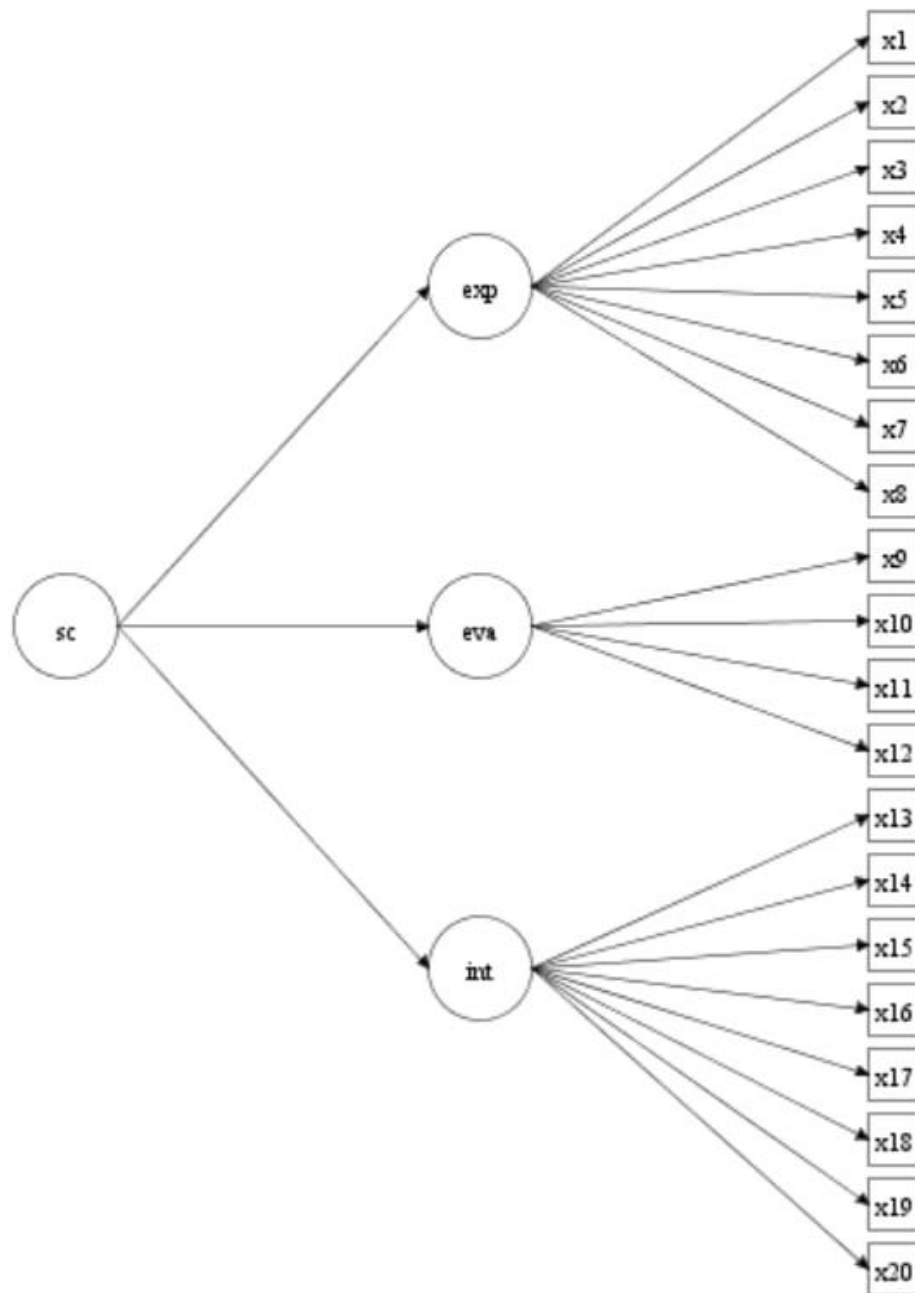
ในการวิจัยครั้งนี้มีการวิเคราะห์ความตรงเชิงโครงสร้างของแบบสอบ โดยใช้วิธีวิเคราะห์องค์ประกอบเชิงยืนยัน (Confirmatory Factor Analysis) ในการพิจารณาความสอดคล้องระหว่างแบบสอบกับคุณลักษณะที่ต้องการวัด โดยการประมาณค่าพารามิเตอร์ของโมเดลด้วยโปรแกรมสำเร็จรูป Mplus และเป็น การประมาณค่าความน่าจะเป็นสูงสุด (Maximum Likelihood) การตรวจสอบความสอดคล้องระหว่างโมเดลและข้อมูล ทำได้โดยการพิจารณาดัชนี 5 ค่า ได้แก่ ค่าไคสแควร์ (χ^2), ไคสแควร์ต่อองศาอิสระ (χ^2/df), RMSEA, CFI และ SRMR ระหว่างวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่แตกต่างกันคือ ความมั่นใจ 2 ระดับ และความมั่นใจ 4 ระดับ เพื่อเป็นการเปรียบเทียบความตรงเชิงโครงสร้างระหว่างวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่แตกต่างกันคือ ความมั่นใจ 2 ระดับ และความมั่นใจ 4 ระดับ โดยค่าดัชนีต้องผ่านเกณฑ์อย่างน้อย 3 ดัชนี

โครงสร้างโมเดลของแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่แตกต่างกัน ประกอบด้วยสมรรถนะทางวิทยาศาสตร์ 3 องค์ประกอบ ดังนั้นคุณลักษณะที่ต้องการวัด เท่ากับ 3 ตัวแปร ได้แก่ 1) การอธิบายปรากฏการณ์ในเชิงวิทยาศาสตร์ 2) การประเมินและออกแบบกระบวนการสืบเสาะหาความรู้ทางวิทยาศาสตร์ และ 3) การแปล

ความหมายข้อมูลและการใช้ประจักษ์พยานในเชิงวิทยาศาสตร์ และตัวแปรสังเกตได้เป็นข้อสอบจำนวน 20 ข้อ ($X_1 - X_{20}$) โดยมีสัญลักษณ์ในโมเดล ดังนี้

sc	แทน	สมรรถนะทางวิทยาศาสตร์
exp	แทน	การอธิบายปรากฏการณ์ในเชิงวิทยาศาสตร์
eva	แทน	การประเมินและออกแบบกระบวนการสืบเสาะหาความรู้ทางวิทยาศาสตร์
int	แทน	การแปลความหมายข้อมูลและการใช้ประจักษ์พยานในเชิงวิทยาศาสตร์
X_1-X_8	แทน	คะแนนที่ได้จากการวัดการอธิบายปรากฏการณ์ในเชิงวิทยาศาสตร์
X_9-X_{12}	แทน	คะแนนที่ได้จากการวัดการประเมินและออกแบบกระบวนการสืบเสาะหาความรู้ทางวิทยาศาสตร์
$X_{13}-X_{20}$	แทน	คะแนนที่ได้จากการวัดการแปลความหมายข้อมูลและการใช้ประจักษ์พยานในเชิงวิทยาศาสตร์

โมเดลของแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่แตกต่างกันสำหรับการวิเคราะห์องค์ประกอบเชิงยืนยันของสมรรถนะทางวิทยาศาสตร์ที่มีคุณลักษณะที่ต้องการวัด 3 ตัวแปร และตัวแปรสังเกตได้ 20 ตัวแปร แสดงดังภาพ 8



ภาพ 8 โครงสร้างของแบบสอบสำหรับการวิเคราะห์องค์ประกอบเชิงยืนยัน

ในการตรวจสอบความตรงเชิงโครงสร้างของแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่แตกต่างกัน โดยใช้วิธีวิเคราะห์องค์ประกอบเชิงยืนยันของงานวิจัยนี้ ประกอบด้วยการวิเคราะห์ความสอดคล้องระหว่างแบบสอบกับคุณลักษณะที่ต้องการวัดทั้งหมด 2 โมเดล ดังนี้

- 1) โมเดลที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 2 ระดับ
- 2) โมเดลที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 4 ระดับ

2.1 ผลการตรวจสอบความสอดคล้องระหว่างโมเดลกับข้อมูล

การวิเคราะห์ความตรงเชิงโครงสร้างของแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่แตกต่างกัน ด้วยโปรแกรมสำเร็จรูป Mplus โดยใช้วิธีการประมาณค่าแบบ Maximum Likelihood โดยพิจารณาจากค่า p-value ที่มากกว่า 0.05 ค่า CFI เข้าใกล้ 1 และค่า RMSEA และ SRMR เข้าใกล้ 0 ผลการตรวจสอบความสอดคล้องระหว่างโมเดลกับข้อมูลทั้ง 2 โมเดล พบว่า โมเดลที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 2 ระดับ มีดัชนีความสอดคล้องกับข้อมูล ดังนี้ $\chi^2 = 131.40$, p-value = 0.0624, CFI = 0.99, RMSEA = 0.02 และ SRMR = 0.03 และโมเดลที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 4 ระดับ มีดัชนีความสอดคล้องกับข้อมูล ดังนี้ $\chi^2 = 189.26$, p-value = 0.0509, CFI = 0.94, RMSEA = 0.02 และ SRMR = 0.04 แสดงว่า ทั้ง 2 โมเดลมีความตรงเชิงโครงสร้าง โดยโมเดลที่ดีที่สุดคือ โมเดลที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 2 ระดับ เนื่องจากมีค่า AIC และ BIC ต่ำสุด ผลการวิเคราะห์ข้อมูล แสดงดังตาราง 17

ตาราง 17 ผลการวิเคราะห์ความสอดคล้องระหว่างโมเดลกับข้อมูลที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่แตกต่างกัน

Model Tested	χ^2	p	df	χ^2/df	RMSEA	CFI	SRMR	AIC	BIC
ระดับความ มั่นใจ 2 ระดับ	131.40	0.0624	108	1.22	0.02	0.99	0.03	16755.05	17234.14
ระดับความ มั่นใจ 4 ระดับ	189.26	0.0509	159	1.19	0.02	0.94	0.04	24714.63	24993.44

2.2 ผลการวิเคราะห์น้ำหนักองค์ประกอบของแต่ละโมเดล

การวิเคราะห์น้ำหนักองค์ประกอบของแต่ละโมเดลด้วยโปรแกรมสำเร็จรูป Mplus โดยใช้วิธีการประมาณค่าแบบ Maximum Likelihood โดยน้ำหนักองค์ประกอบอันดับที่ 1 และอันดับที่ 2 ควรมีค่ามากกว่า 0.30 (สุวิมล ติรกานันท์, 2555)

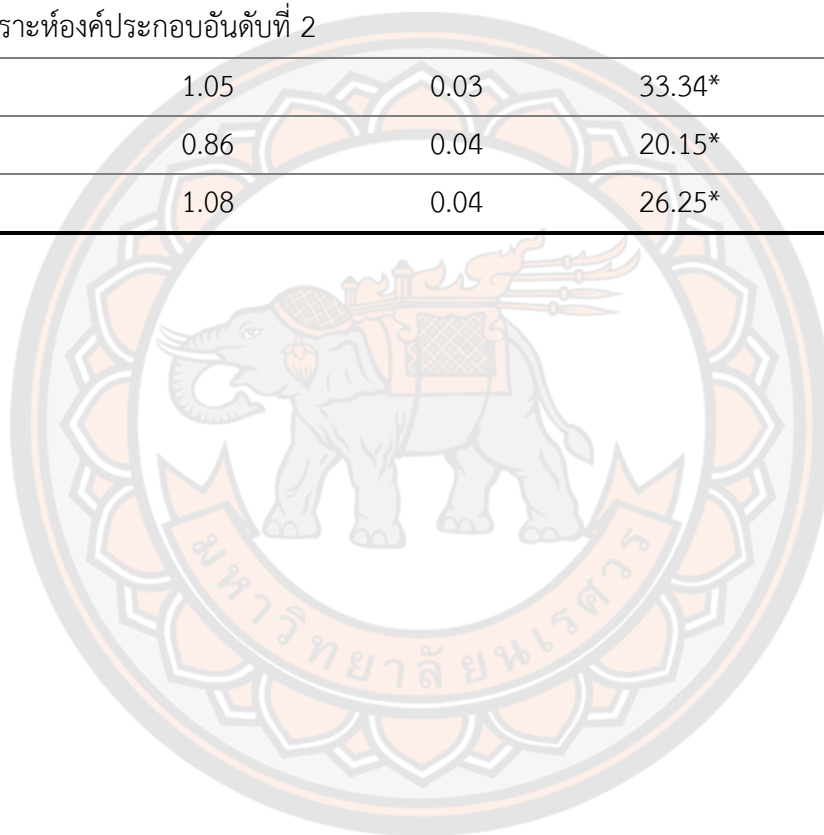
2.2.1 โมเดลที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 2 ระดับ

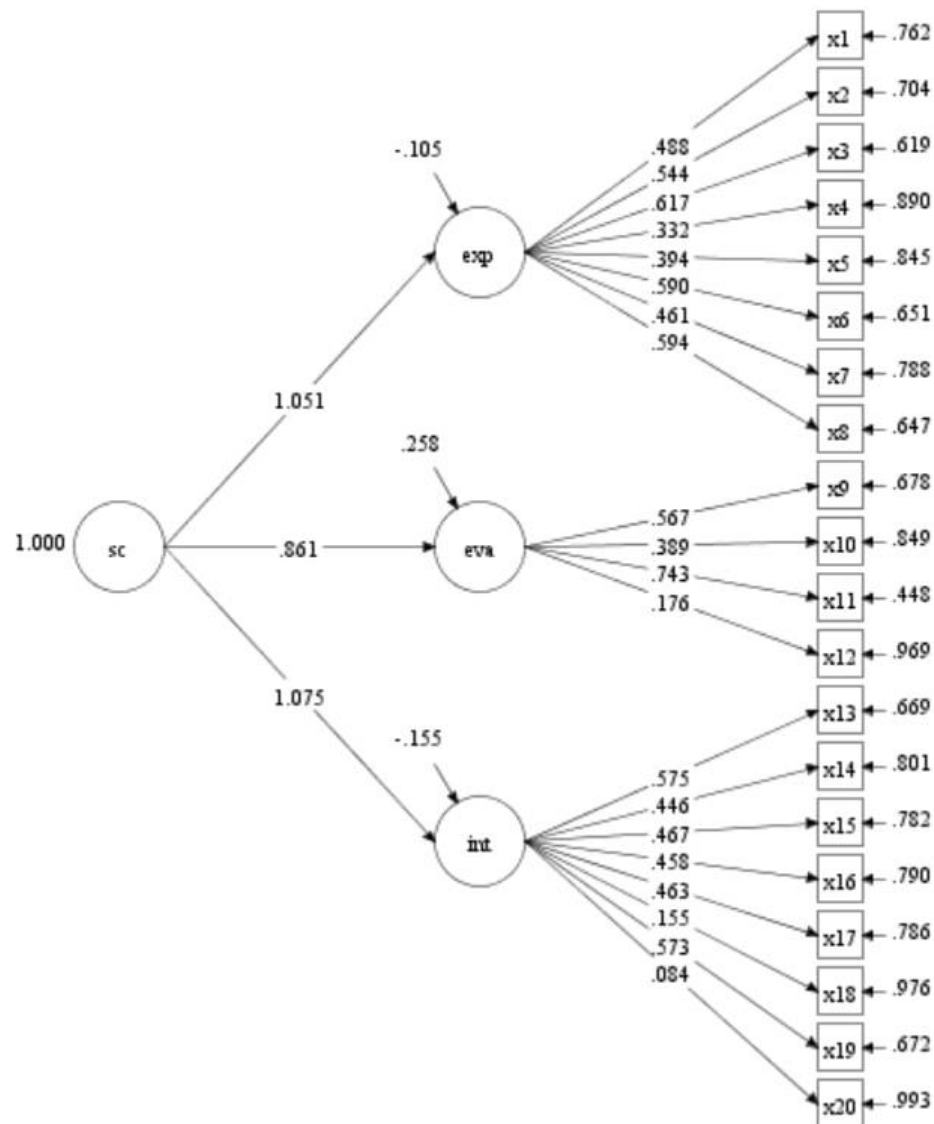
จากการวิเคราะห์องค์ประกอบอันดับที่ 1 ของโมเดลกับข้อมูลที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 2 ระดับ พบว่า ข้อคำถามมีน้ำหนักองค์ประกอบอยู่ระหว่าง 0.08 ถึง 0.74 โดยข้อคำถามส่วนใหญ่มีน้ำหนักองค์ประกอบมากกว่า 0.30 ยกเว้น ข้อ 12, ข้อ 18 และข้อ 20 และมีความสัมพันธ์ระหว่างข้อคำถามกับองค์ประกอบอันดับที่ 1 อยู่ระหว่างร้อยละ 1 ถึง 55 ซึ่งอยู่ในระดับปานกลาง ส่วนการวิเคราะห์องค์ประกอบอันดับที่ 2 พบว่า ข้อคำถามมีน้ำหนักองค์ประกอบอยู่ระหว่าง 0.86 ถึง 1.08 โดยคุณลักษณะที่ต้องการวัดทั้ง 3 องค์ประกอบ มีน้ำหนักองค์ประกอบมากกว่า 0.30 และมีความสัมพันธ์ระหว่างคุณลักษณะกับองค์ประกอบอันดับที่ 2 เท่ากับร้อยละ 74 ซึ่งอยู่ในระดับสูง ผลการวิเคราะห์ข้อมูล แสดงดังตาราง 18

ตาราง 18 ผลการวิเคราะห์น้ำหนักองค์ประกอบของโมเดลที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 2 ระดับ

ตัวแปร	น้ำหนักองค์ประกอบ	S.E.	<i>t</i>	<i>R</i> ²
การวิเคราะห์องค์ประกอบอันดับที่ 1				
X ₁	0.49	0.04	11.56*	0.24*
X ₂	0.54	0.04	13.51*	0.30*
X ₃	0.62	0.04	16.64*	0.38*
X ₄	0.33	0.05	6.38*	0.11*
X ₅	0.39	0.04	8.96*	0.16*
X ₆	0.59	0.04	16.56*	0.35*
X ₇	0.46	0.04	10.67*	0.21*
X ₈	0.59	0.04	15.96*	0.35*
X ₉	0.57	0.04	13.42*	0.32*
X ₁₀	0.39	0.05	7.76*	0.15*
X ₁₁	0.74	0.04	18.30*	0.55*
X ₁₂	0.18	0.06	3.22*	0.03
X ₁₃	0.58	0.04	14.68*	0.33*
X ₁₄	0.45	0.05	9.31*	0.20*

ตัวแปร	น้ำหนักองค์ประกอบ	S.E.	<i>t</i>	<i>R</i> ²
X ₁₅	0.47	0.04	11.07*	0.22*
X ₁₆	0.46	0.04	11.23*	0.21*
X ₁₇	0.46	0.05	10.39*	0.21*
X ₁₈	0.16	0.05	3.02*	0.02
X ₁₉	0.57	0.04	13.55*	0.33*
X ₂₀	0.08	0.05	1.61	0.01
การวิเคราะห์องค์ประกอบอันดับที่ 2				
exp	1.05	0.03	33.34*	-
eva	0.86	0.04	20.15*	0.74*
int	1.08	0.04	26.25*	-





ภาพ 9 โมเดลที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 2 ระดับ

2.2.2 โมเดลที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 4 ระดับ

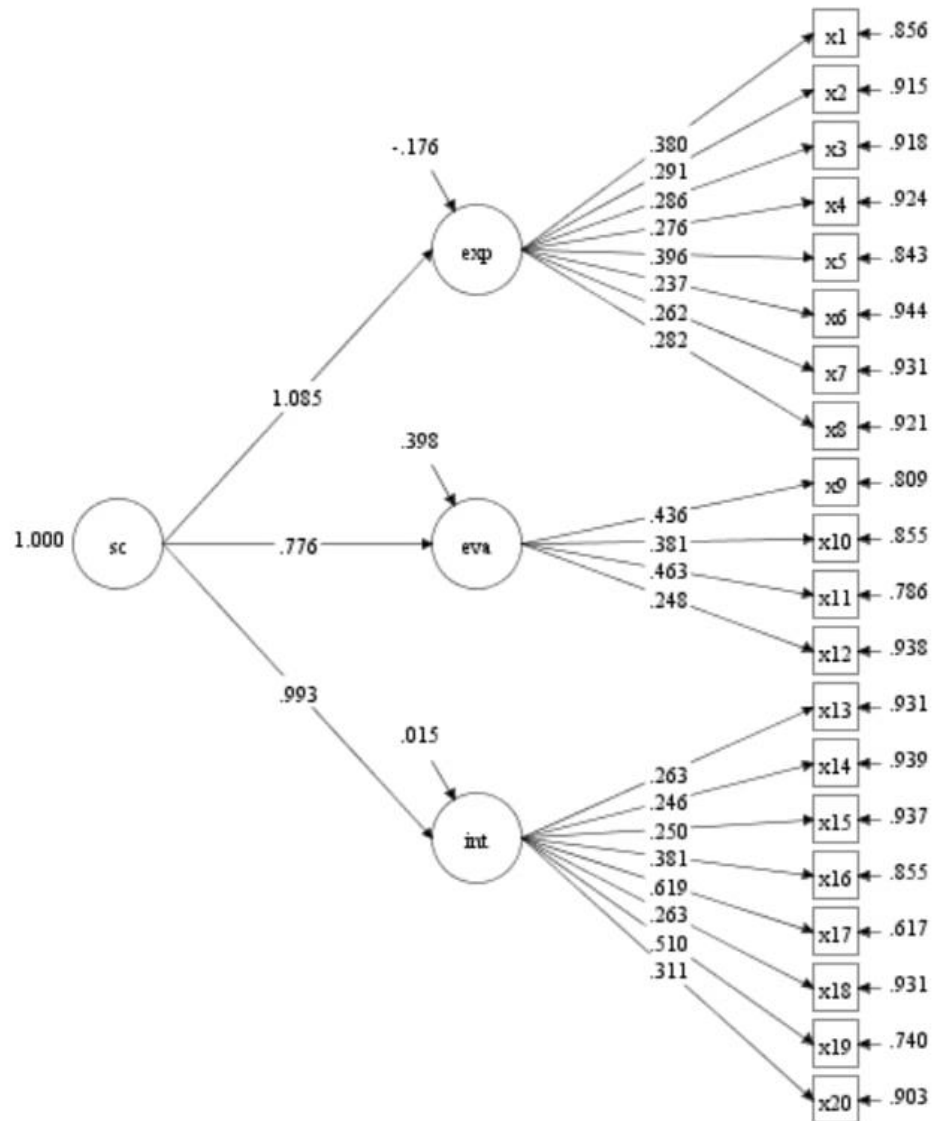
จากการวิเคราะห์องค์ประกอบอันดับที่ 1 ของโมเดลกับข้อมูลที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 4 ระดับ พบว่า ข้อคำถามมีน้ำหนักองค์ประกอบอยู่ระหว่าง 0.24 ถึง 0.62 โดยมีข้อคำถามที่มีน้ำหนักองค์ประกอบน้อยกว่า 0.30 ได้แก่ ข้อ 2, ข้อ 3, ข้อ 4, ข้อ 6, ข้อ 7, ข้อ 8, ข้อ 12, ข้อ 13, ข้อ 14, ข้อ 15 และข้อ 18 และมีความสัมพันธ์ระหว่างข้อคำถามกับองค์ประกอบอันดับที่ 1 อยู่ระหว่างร้อยละ 6 ถึง 38 ซึ่งอยู่ในระดับ

ค่อนข้างต่ำ ส่วนการวิเคราะห์องค์ประกอบอันดับที่ 2 พบว่า ข้อคำถามมีน้ำหนักองค์ประกอบอยู่ระหว่าง 0.78 ถึง 1.09 โดยคุณลักษณะที่ต้องการวัดทั้ง 3 องค์ประกอบ มีน้ำหนักองค์ประกอบมากกว่า 0.30 และมีความสัมพันธ์ระหว่างคุณลักษณะกับองค์ประกอบอันดับที่ 2 อยู่ระหว่างร้อยละ 60 ถึง 99 ซึ่งอยู่ในระดับสูง ผลการวิเคราะห์ข้อมูล แสดงดังตาราง 19

ตาราง 19 ผลการวิเคราะห์น้ำหนักองค์ประกอบของโมเดลที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 4 ระดับ

ตัวแปร	น้ำหนักองค์ประกอบ	S.E.	t	R ²
การวิเคราะห์องค์ประกอบอันดับที่ 1				
X ₁	0.38	0.05	6.97*	0.14*
X ₂	0.29	0.06	5.32*	0.09*
X ₃	0.29	0.06	5.25*	0.08*
X ₄	0.28	0.05	5.07*	0.08*
X ₅	0.40	0.05	7.46*	0.16*
X ₆	0.24	0.06	4.15*	0.06*
X ₇	0.26	0.06	4.77*	0.07*
X ₈	0.28	0.06	5.14*	0.08*
X ₉	0.44	0.06	7.40*	0.19*
X ₁₀	0.38	0.07	5.41*	0.15*
X ₁₁	0.46	0.07	7.12*	0.21*
X ₁₂	0.25	0.07	3.70*	0.06
X ₁₃	0.26	0.06	4.67*	0.07*
X ₁₄	0.25	0.06	4.38*	0.06*
X ₁₅	0.25	0.06	4.44*	0.06*
X ₁₆	0.38	0.05	7.27*	0.15*
X ₁₇	0.62	0.05	13.80*	0.38*
X ₁₈	0.26	0.06	4.71*	0.07*
X ₁₉	0.51	0.05	10.57*	0.26*
X ₂₀	0.31	0.06	5.60*	0.10*
การวิเคราะห์องค์ประกอบอันดับที่ 2				

ตัวแปร	น้ำหนักองค์ประกอบ	S.E.	<i>t</i>	<i>R</i> ²
exp	1.09	0.08	13.16*	-
eva	0.78	0.08	9.79*	0.60*
int	0.99	0.07	15.01*	0.99*



ภาพ 10 โมเดลที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 4 ระดับ

ตอนที่ 3 ผลการวิเคราะห์ความเที่ยงของแบบสอบ

การวิเคราะห์ความเที่ยงของแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่ต่างกันคือ ความมั่นใจ 2 ระดับ และความมั่นใจ 4 ระดับ โดยใช้ความเที่ยงแบบความสอดคล้องภายใน วิธีสัมประสิทธิ์แอลฟาครอนบาค (Cronbach's Alpha) พบว่า แบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 2 ระดับ มีค่าความเที่ยงสูงกว่า แบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 4 ระดับ มีค่าความเที่ยงเท่ากับ 0.84 และ 0.70 ตามลำดับ ผลการวิเคราะห์ข้อมูลแสดงดังตาราง 20

ตาราง 20 ผลการวิเคราะห์ความเที่ยงของแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่ต่างกัน

วิธีการตรวจให้คะแนน	Cronbach's Alpha
ระดับความมั่นใจ 2 ระดับ	0.84
ระดับความมั่นใจ 4 ระดับ	0.70

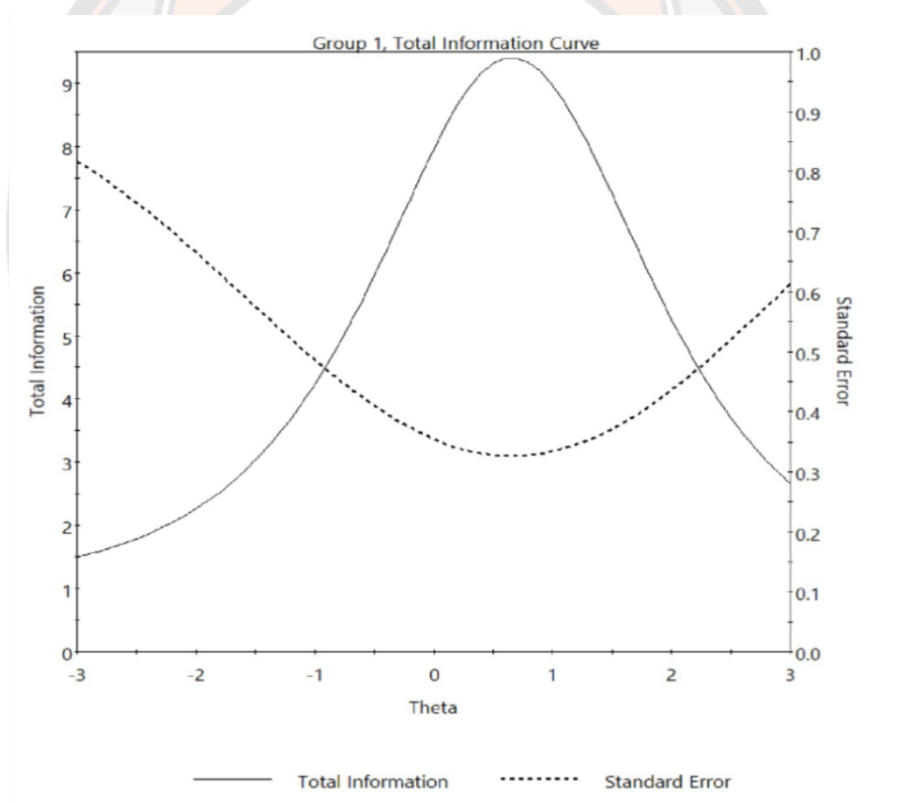
ตอนที่ 4 ผลการวิเคราะห์ฟังก์ชันสารสนเทศของแบบสอบ

การวิเคราะห์ฟังก์ชันสารสนเทศของแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่ต่างกัน ด้วยโปรแกรมสำเร็จรูป IRTPRO4 Student เพื่ออธิบายความสัมพันธ์ระหว่างความสามารถของผู้ตอบและโอกาสในการเลือกตอบแต่ละรายการที่กำหนด ซึ่งวิเคราะห์ด้วย Generalized Partial Credit Model (G-PCM)

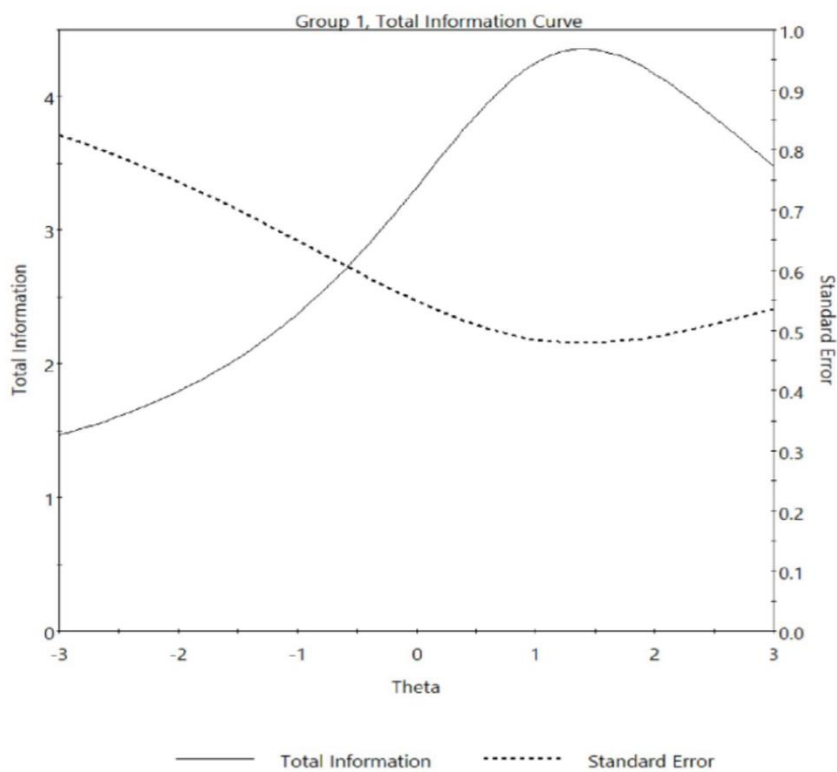
การวิเคราะห์ฟังก์ชันสารสนเทศของแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 2 ระดับ พบว่า วิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 2 ระดับ ให้สารสนเทศสูงเมื่อผู้สอบมีความสามารถระดับสูง (θ มีค่าระหว่าง 0.4 ถึง 0.8) และวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 4 ระดับ ให้สารสนเทศสูงเมื่อผู้สอบมีความสามารถระดับสูง (θ มีค่าระหว่าง 1.2 ถึง 1.6) ผลการวิเคราะห์ข้อมูลแสดงดังตาราง 21

ตาราง 21 ผลการวิเคราะห์ฟังก์ชันสารสนเทศของแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่แตกต่างกัน

วิธีการตรวจ	ระดับความสามารถของผู้สอบ (θ)														
ให้คะแนน	-2.8	-2.4	-2.0	-1.6	-1.2	-0.8	-0.4	-0.0	0.4	0.8	1.2	1.6	2.0	2.4	2.8
ระดับความ มั่นใจ 2	1.60	1.86	2.26	2.86	3.69	4.86	6.35	7.96	9.15	9.32	8.39	6.84	5.27	3.97	3.02
ระดับ ระดับความ มั่นใจ 4	1.52	1.64	1.79	1.99	2.23	2.53	2.90	3.32	3.76	4.12	4.33	4.33	4.17	3.91	3.63



ภาพ 11 โค้งฟังก์ชันสารสนเทศของแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 2 ระดับ



ภาพ 12 โค้งฟังก์ชันสารสนเทศของแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 4 ระดับ

ตอนที่ 5 ผลการเปรียบเทียบคุณภาพของแบบสอบ

การเปรียบเทียบคุณภาพของแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่แตกต่างกัน พิจารณาจากการตรวจสอบข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ ความยากและอำนาจจำแนกเฉลี่ย ความตรงเชิงโครงสร้าง ความเที่ยง และฟังก์ชันสารสนเทศของแบบสอบเฉลี่ย พบว่า แบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 2 ระดับ มีคุณภาพดีกว่าแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 4 ระดับ ผลการเปรียบเทียบคุณภาพของแบบสอบ แสดงดังตาราง 22

ตาราง 22 เปรียบเทียบคุณภาพของแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่แตกต่างกัน

วิธีการตรวจ ให้คะแนน	การเปรียบเทียบคุณภาพ					
	การ ตรวจสอบ ข้อตกลง เบื้องต้น	ความ ยาก เฉลี่ย	อำนาจ จำแนก เฉลี่ย	ความตรงเชิง โครงสร้าง	ความเที่ยง	ฟังก์ชัน สารสนเทศ
ระดับความ มั่นใจ 2 ระดับ	ผ่านทั้ง 2 เกณฑ์	1.32	0.69	มีความตรงเชิง โครงสร้าง และมี น้ำหนัก องค์ประกอบไม่ ผ่านเกณฑ์จำนวน 3 ข้อ	0.84	0.4-0.8
ระดับความ มั่นใจ 4 ระดับ	ผ่าน 1 เกณฑ์	2.72	0.26	มีความตรงเชิง โครงสร้าง และมี น้ำหนัก องค์ประกอบไม่ ผ่านเกณฑ์จำนวน 11 ข้อ	0.70	1.2-1.6

บทที่ 5

บทสรุป

งานวิจัยเรื่อง การเปรียบเทียบคุณภาพของแบบสอบเลือกตอบสองระดับด้านสมรรถนะทางวิทยาศาสตร์ที่มีระดับความมั่นใจที่แตกต่างกัน โดยประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบแบบตรวจให้คะแนนความรู้บางส่วนแบบทั่วไป เป็นงานวิจัยที่มีตัวแปรอิสระคือ แบบสอบเลือกตอบสองระดับที่มีระดับความมั่นใจที่แตกต่างกัน 2 แบบ ได้แก่ ระดับความมั่นใจ 2 ระดับ และระดับความมั่นใจ 4 ระดับ ส่วนตัวแปรตามคือ คุณภาพของวิธีการตรวจให้คะแนนความรู้บางส่วน แบ่งออกเป็น 4 ด้าน ได้แก่ 1) ความยากและอำนาจจำแนกของแบบสอบ 2) ความตรงเชิงโครงสร้างของแบบสอบ 3) ความเที่ยงของแบบสอบ และ 4) ฟังก์ชันสารสนเทศของแบบสอบ โดยมีวัตถุประสงค์การวิจัย 4 ประการ ดังนี้ 1) เพื่อเปรียบเทียบความยากและอำนาจจำแนกของแบบสอบ 2) เพื่อเปรียบเทียบความตรงเชิงโครงสร้างของแบบสอบ 3) เพื่อเปรียบเทียบความเที่ยงของแบบสอบ และ 4) เพื่อเปรียบเทียบฟังก์ชันสารสนเทศของแบบสอบ

ประชากรที่ใช้ในงานวิจัยนี้คือ นักเรียนระดับชั้นมัธยมศึกษาปีที่ 4-6 จำนวน 21,663 คน จำนวน 57 โรงเรียน สังกัดสำนักงานเขตพื้นที่การศึกษามัธยมศึกษาพิษณุโลก อุตรดิตถ์ และกลุ่มตัวอย่างได้จากการสุ่มแบบหลายขั้นตอน (Multi-stage Sampling) จากนักเรียนระดับชั้นมัธยมศึกษาปีที่ 4-6 สังกัดสำนักงานเขตพื้นที่การศึกษามัธยมศึกษาพิษณุโลก อุตรดิตถ์ จำนวน 39 โรงเรียนในเขตจังหวัดพิษณุโลก ได้กลุ่มตัวอย่างจำนวน 750 คน แบ่งเป็นกลุ่มตัวอย่างกลุ่มที่ 1 จำนวน 375 คน และกลุ่มที่ 2 จำนวน 375 คน

เครื่องมือที่ใช้ในงานวิจัยครั้งนี้คือ แบบสอบเลือกตอบสองระดับด้านสมรรถนะทางวิทยาศาสตร์ รายวิชาวิทยาศาสตร์ชีวภาพ ระดับชั้นมัธยมศึกษาตอนปลาย ซึ่งมีลักษณะเป็นแบบสอบเลือกตอบสองระดับคือ ระดับเนื้อหา (Content Tier) และระดับความมั่นใจ (Confidence Tier) จำนวน 2 ฉบับ ฉบับละ 20 ข้อ

การเก็บรวบรวมข้อมูลในงานวิจัยครั้งนี้ ผู้สอบจะได้ทำแบบสอบคนละ 1 ฉบับ โดยแบบสอบฉบับที่ 1 มีระดับความมั่นใจ 2 ระดับ ได้แก่ มั่นใจ และ ไม่มั่นใจ และแบบสอบฉบับที่ 2 มีระดับความมั่นใจ 4 ระดับ ได้แก่ มั่นใจมาก มั่นใจ ไม่มั่นใจ และ เด่า

การวิเคราะห์ข้อมูลในงานวิจัยครั้งนี้ ใช้โปรแกรมสำเร็จรูป IRTPRO4 Student ในการวิเคราะห์ความยาก อำนาจจำแนก และฟังก์ชันสารสนเทศของแบบสอบ ใช้โปรแกรมสำเร็จรูป

Mplus ในการวิเคราะห์ความตรงเชิงโครงสร้าง ใช้โปรแกรมสำเร็จรูป SPSS ในการวิเคราะห์ความเที่ยง และตรวจสอบข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ

สรุปผลการวิจัย

ผู้วิจัยสรุปผลการวิจัยตามวัตถุประสงค์ของการวิจัย โดยแบ่งเป็น 4 ตอน ได้แก่ ตอนที่ 1 ผลการเปรียบเทียบความยากและอำนาจจำแนกของแบบสอบ ตอนที่ 2 ผลการเปรียบเทียบความตรงเชิงโครงสร้างของแบบสอบ ตอนที่ 3 ผลการเปรียบเทียบความเที่ยงของแบบสอบ ตอนที่ 4 ผลการเปรียบเทียบฟังก์ชันสารสนเทศของแบบสอบ และตอนที่ 5 ผลการเปรียบเทียบคุณภาพของแบบสอบ ซึ่งมีรายละเอียดแต่ละตอนดังนี้

ตอนที่ 1 ผลการเปรียบเทียบความยากและอำนาจจำแนกของแบบสอบ

การตรวจสอบข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ โดยการตรวจสอบความเป็นเอกมิติของแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่แตกต่างกัน โดยใช้การวิเคราะห์องค์ประกอบเชิงยืนยัน ซึ่งสามารถบ่งบอกความเป็นเอกมิติของแบบสอบได้ทั้ง 2 ฉบับ จึงสามารถนำไปวิเคราะห์สารสนเทศของแบบสอบด้วยทฤษฎีการตอบสนองข้อสอบได้

การวิเคราะห์ความยากและอำนาจจำแนกของแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่แตกต่างกัน ด้วยโปรแกรมสำเร็จรูป IRTPRO4 Student และวิเคราะห์ด้วย Generalized Partial Credit Model (G-PCM) พบว่า แบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 2 ระดับ และ 4 ระดับ เป็นแบบสอบที่ค่อนข้างยากและอำนาจจำแนกค่อนข้างต่ำ เมื่อวิเคราะห์ความแตกต่างระหว่างค่าความยากและอำนาจจำแนกเฉลี่ยของแบบสอบ พบว่า วิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่แตกต่างกันไม่ส่งผลต่อความยากของแบบสอบ อย่างมีนัยสำคัญทางสถิติที่ ระดับ .05 แต่วิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่แตกต่างกันส่งผลต่ออำนาจจำแนก อย่างมีนัยสำคัญทางสถิติที่ ระดับ .05

ตอนที่ 2 ผลการเปรียบเทียบความตรงเชิงโครงสร้างของแบบสอบ

การตรวจสอบความสอดคล้องระหว่างโมเดลกับข้อมูล พบว่า โมเดลทั้ง 2 โมเดล ได้แก่ โมเดลที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 2 ระดับ และ โมเดลที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 4 ระดับ มีความสอดคล้องระหว่างโมเดลกับข้อมูล หมายความว่า ทั้ง 2 โมเดลนี้มีความตรงเชิงโครงสร้าง

ตอนที่ 3 ผลการเปรียบเทียบความเที่ยงของแบบสอบ

การวิเคราะห์ความเที่ยงของแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่แตกต่างกัน โดยใช้ความเที่ยงแบบความสอดคล้องภายใน วิธีสัมประสิทธิ์แอลฟาครอนบาค (Cronbach's Alpha) พบว่า วิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 2 ระดับ มีค่าความเที่ยงสูงกว่าวิธีการให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 4 ระดับ

ตอนที่ 4 ผลการเปรียบเทียบฟังก์ชันสารสนเทศของแบบสอบ

การวิเคราะห์ฟังก์ชันสารสนเทศของแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่แตกต่างกัน พบว่า วิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 2 ระดับ และระดับความมั่นใจ 4 ระดับ ให้สารสนเทศสูงเมื่อผู้สอบมีความสามารถระดับสูง

ตอนที่ 5 ผลการเปรียบเทียบคุณภาพของแบบสอบ

การเปรียบเทียบคุณภาพของแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่แตกต่างกัน พบว่า แบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 2 ระดับ มีคุณภาพดีกว่าแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 4 ระดับ

อภิปรายผลการวิจัย

จากการดำเนินการวิจัยและผลการวิจัย มีประเด็นที่น่าสนใจเกี่ยวกับการเปรียบเทียบความยากและอำนาจจำแนกของแบบสอบ ความตรงเชิงโครงสร้างของแบบสอบ ความเที่ยงของแบบสอบ และฟังก์ชันสารสนเทศของแบบสอบ โดยมีรายละเอียดดังนี้

1. การเปรียบเทียบความยากและอำนาจจำแนกของแบบสอบ

การวิเคราะห์ความยากและอำนาจจำแนกของแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่แตกต่างกัน โดยพิจารณาค่าความยาก (b) ควรอยู่ระหว่าง -4.0 ถึง +4.0 และค่าอำนาจจำแนก (a) ควรอยู่ระหว่าง 0 ถึง ∞ หมายความว่า ข้อสอบสามารถนำไปใช้ได้ (ณภัทร ชัยมงคล, 2559) จากข้อสอบทั้ง 2 ฉบับ ฉบับละ 20 ข้อ พบว่า แบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 2 ระดับ อยู่ในเกณฑ์ที่กำหนดจำนวน 19 ข้อ คิดเป็นร้อยละ 95 โดยมีค่าความยากเฉลี่ยเท่ากับ 1.32 และมีค่าอำนาจ

จำแนกเฉลี่ยเท่ากับ 0.69 แสดงว่า ข้อสอบค่อนข้างยาก และอำนาจจำแนกค่อนข้างต่ำ ส่วนแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 4 ระดับ อยู่ในเกณฑ์ที่กำหนดจำนวน 19 ข้อ คิดเป็นร้อยละ 95 โดยมีค่าความยากเฉลี่ยเท่ากับ 2.72 และค่าอำนาจจำแนกเฉลี่ยเท่ากับ 0.26 แสดงว่า ข้อสอบค่อนข้างยาก และอำนาจจำแนกค่อนข้างต่ำ

การวิเคราะห์ค่าความยากของแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่แตกต่างกัน พบว่า วิธีการให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 2 ระดับ และ 4 ระดับ มีค่าความยากแตกต่างกัน อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ซึ่งสอดคล้องกับงานวิจัยของณัฐภรณ์ เลขะวัฒนพงษ์ (2560) รัตนา ไชยตรี (2546) และเอมอร จังศิริพรภรณ์ (2545) เนื่องจากแต่ละวิธีมีรูปแบบการตอบและการตรวจให้คะแนนที่แตกต่างกัน ส่งผลต่อการพิจารณาเลือกคำตอบ คือเมื่อตอบระดับความมั่นใจที่สูงขึ้น คะแนนก็จะเพิ่มมากขึ้นตามไปด้วย สอดคล้องกับงานวิจัยของ Bahrami (2011) ที่มีการศึกษาข้อสอบที่มีจำนวนตัวเลือกที่แตกต่างกัน ทำให้ส่งผลต่อความยากของข้อสอบอย่างชัดเจน

การวิเคราะห์ค่าอำนาจจำแนกของแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่แตกต่างกัน พบว่า วิธีการให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 2 ระดับ และ 4 ระดับ มีค่าอำนาจจำแนกแตกต่างกัน อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 การที่วิธีการตรวจให้คะแนน เมื่อข้อสอบมีระดับความมั่นใจ 2 ระดับ มีค่าอำนาจจำแนกสูงกว่าอาจเกิดจากกรณีที่ผู้สอบที่มั่นใจและไม่มั่นใจในคำตอบสามารถแสดงความมั่นใจที่แท้จริงออกมาได้ เนื่องจากวิธีการตอบมีเพียง 2 ระดับ นั่นคือ “มั่นใจ” และ “ไม่มั่นใจ” ซึ่งต่างจากวิธีการตรวจให้คะแนน เมื่อข้อสอบมีระดับความมั่นใจ 4 ระดับ ที่ผู้ตอบไม่สามารถแสดงความมั่นใจที่แท้จริงออกมาได้ ผู้สอบส่วนใหญ่มักจะตอบระดับความมั่นใจเป็น “มั่นใจ” หรือ “ไม่มั่นใจ” มากกว่าจะตอบระดับความมั่นใจเป็น “มั่นใจมาก” หรือ “เดา”

2. การเปรียบเทียบความตรงเชิงโครงสร้างของแบบสอบ

การตรวจสอบความสอดคล้องระหว่างโมเดลกับข้อมูล พบว่า โมเดลทั้ง 2 โมเดล ได้แก่ โมเดลที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 2 ระดับ และ โมเดลที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 4 ระดับ มีความตรงเชิงโครงสร้าง ซึ่งแต่ละวิธีมีค่าความตรงเชิงโครงสร้างต่างกัน ซึ่งสอดคล้องกับงานวิจัยของ สุพจน์ เกิดสุวรรณ (2545) ประกอบกับเครื่องมือที่ใช้ในการวิจัยมีการตรวจสอบความตรงเชิงเนื้อหา ดังนั้นจึงทำให้โมเดลที่วิเคราะห์มีความตรงเชิงโครงสร้างด้วยเช่นกัน

เมื่อพิจารณาน้ำหนักองค์ประกอบของแต่ละข้อ พบว่า ข้อที่มีน้ำหนักองค์ประกอบน้อยกว่า 0.30 ได้แก่ ข้อ 12, ข้อ 18 และ ข้อ 20 เนื่องจากเป็นข้อสอบที่มีรายละเอียดมาก และผู้สอบส่วน

ใหญ่มีความเข้าใจคลาดเคลื่อน ทำให้ผู้สอบมีโอกาสผิดมากกว่าข้ออื่น ๆ ดังนั้นผู้สอบที่มีความสามารถระดับสูง ระดับปานกลาง และระดับต่ำ อาจจะทำข้อสอบข้อนี้ผิดเหมือนกัน ทำให้มีค่าน้ำหนักองค์ประกอบต่ำ

3. การเปรียบเทียบความเที่ยงของแบบสอบ

การวิเคราะห์ความเที่ยงของแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่แตกต่างกัน พบว่า ค่าความเที่ยงของทุกวิธีมีค่าสูงกว่า 0.70 ซึ่งอยู่ในเกณฑ์ที่ยอมรับได้ (ศิริชัย กาญจนวาสี, 2555) โดยวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 2 ระดับ มีความเที่ยงสูงกว่าวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 4 ระดับ เนื่องจาก วิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 2 ระดับ มีช่วงคะแนนที่แคบ ทำให้การตอบระดับความมั่นใจตรงกับคุณลักษณะของผู้ตอบมากกว่า ผู้สอบไม่จำเป็นต้องเสี่ยงในการตอบข้อสอบ ส่งผลให้ความแปรปรวนของคะแนนสอบต่ำ ดังนั้นจึงทำให้มีความเที่ยงสูงกว่า

4. การเปรียบเทียบฟังก์ชันสารสนเทศของแบบสอบ

การวิเคราะห์ฟังก์ชันสารสนเทศของแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่แตกต่างกัน พบว่า วิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 2 ระดับ และ 4 ระดับ ต่างก็ให้สารสนเทศสูงที่สุดในช่วงระดับความสามารถระดับสูง (θ มีค่าระหว่าง 0.4 ถึง 0.8 และ มีค่าระหว่าง 1.2 ถึง 1.6 ตามลำดับ) ซึ่งสอดคล้องกับงานวิจัยของ ญัฐภรณ์ เลขะวัฒนพงษ์ (2560) เนื่องจากผู้สอบที่มีความสามารถระดับสูงมีแนวโน้มที่จะทำข้อสอบได้ถูกต้องอย่างสมบูรณ์โดยไม่ขึ้นกับวิธีการตอบและการตรวจให้คะแนนความรู้บางส่วน นอกจากนี้เมื่อพิจารณาในทุกช่วงระดับความสามารถ พบว่า สารสนเทศของแบบสอบแตกต่างกันมาก เมื่อผู้สอบมีความสามารถระดับปานกลาง และสารสนเทศของแบบสอบไม่แตกต่างกันเมื่อผู้สอบมีความสามารถระดับต่ำและสูง โดยผู้สอบที่มีความสามารถระดับต่ำส่วนใหญ่จะเดา ทำให้ผู้สอบมีโอกาสได้ 0 คะแนนมากขึ้น ส่วนผู้สอบที่มีความสามารถระดับสูงส่วนใหญ่จะทำข้อสอบได้ถูกต้องสมบูรณ์โดยไม่ขึ้นอยู่กับวิธีการตอบและการตรวจให้คะแนน เนื่องจากผู้สอบที่มีความสามารถระดับสูงมีความรู้มากพอที่จะตอบได้อย่างมั่นใจ

5. การเปรียบเทียบคุณภาพของแบบสอบ

การเปรียบเทียบคุณภาพของแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่แตกต่างกัน โดยพิจารณาจากการตรวจสอบข้อตกลงเบื้องต้นของทฤษฎี

การตอบสนองข้อสอบ ความยากและอำนาจจำแนกเฉลี่ย ความตรงเชิงโครงสร้าง ความเที่ยง และ ฟังก์ชันสารสนเทศของแบบสอบเฉลี่ย พบว่า แบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 2 ระดับ มีคุณภาพดีกว่าแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 4 ระดับ โดยปัจจัยที่คาดว่าจะส่งผลต่อผลการเปรียบเทียบดังกล่าว มาจากค่าที่ได้จากการพิจารณาองค์ประกอบดังนี้ 1) การตรวจสอบข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ พบว่า แบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 4 ระดับ ไม่ผ่านเกณฑ์ด้านอัตราส่วนของค่าไอเกน (มีค่าต่ำกว่า 3) แต่แบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 2 ระดับ ผ่านเกณฑ์การพิจารณาทั้ง 2 เกณฑ์ 2) ความยากและอำนาจจำแนกเฉลี่ยของแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 2 ระดับ มีค่าความยากเฉลี่ยที่เหมาะสม รวมทั้งยังสามารถจำแนกผู้สอบได้ดีกว่าแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 4 ระดับ 3) ความตรงเชิงโครงสร้างของแบบสอบทั้ง 2 ฉบับ มีความตรงเชิงโครงสร้าง แต่เมื่อพิจารณาน้ำหนักองค์ประกอบของแบบสอบแต่ละฉบับ พบว่า แบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 4 ระดับ มีข้อสอบที่ไม่ผ่านเกณฑ์น้ำหนักองค์ประกอบมากกว่าแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 2 ระดับ 4) ความเที่ยงของแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 2 ระดับ มีค่ามากกว่าแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 4 ระดับ และ 5) ฟังก์ชันสารสนเทศของแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 2 ระดับ สามารถสะท้อนความสามารถที่แท้จริงของผู้สอบได้มากกว่าแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 4 ระดับ ดังเหตุผลที่กล่าวมาข้างต้นจึงสนับสนุนว่าแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 2 ระดับมีคุณภาพของแบบสอบที่ดีกว่าแบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจ 4 ระดับ

ข้อเสนอแนะในการนำงานวิจัยไปใช้

1. จากข้อค้นพบในงานวิจัยครั้งนี้ พบว่า แบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่แตกต่างกัน จะให้สารสนเทศสูงสำหรับผู้สอบที่มีระดับความสามารถสูง ดังนั้น แบบสอบที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน เมื่อข้อสอบมีระดับความมั่นใจที่แตกต่างกันเหมาะสำหรับการนำไปใช้กับผู้สอบที่มีระดับความสามารถสูง

2. จากข้อค้นพบในงานวิจัยครั้งนี้ พบว่า แบบสอบเลือกตอบสองระดับด้านสมรรถนะทางวิทยาศาสตร์ มีส่วนประกอบของระดับความมั่นใจที่เป็นการประเมินค่าภายในของแต่ละบุคคล ดังนั้น ส่วนประกอบของระดับความมั่นใจจึงมีผลต่อคะแนนของผู้สอบ เพื่อให้คะแนนจากแบบสอบนั้นตรงกับความเป็นจริงของผู้สอบ ผู้สอบควรทำแบบสอบด้วยความตั้งใจ และครูผู้ควบคุมการสอบควรชี้แจงรายละเอียดในการทำแบบสอบ

ข้อเสนอแนะในการวิจัยครั้งต่อไป

1. ในการวิจัยครั้งต่อไป นำแนวทางในการเปรียบเทียบคุณภาพของแบบสอบไปใช้เป็นพัฒนาแบบสอบวินิจฉัยมนทัศน์อื่น ๆ ของผู้สอบ โดยใช้รูปแบบของระบบคอมพิวเตอร์แบบปรับเหมาะ ซึ่งอาจจะทำให้สามารถวินิจฉัยนักเรียนรายบุคคลได้โดยใช้ข้อสอบที่มีความเหมาะสมกับผู้สอบเป็นรายบุคคล และทำให้ได้สารสนเทศที่ตรงกับความเป็นจริงของผู้สอบมากขึ้น

2. ในการวิจัยครั้งต่อไป นำแนวทางในการพัฒนารูปแบบของแบบสอบโดยการเพิ่มระดับการตอบจาก 2 ระดับเป็น 3 ระดับหรือ 4 ระดับ โดยระดับแรกคือระดับเนื้อหา ระดับที่สองคือระดับความมั่นใจในการตอบระดับเนื้อหา ระดับที่สามคือระดับเหตุผล และระดับที่สี่คือระดับความมั่นใจในการตอบระดับเหตุผล เพื่อให้ได้สารสนเทศที่เพิ่มขึ้น

บรรณานุกรม



บรรณานุกรม

ภาษาไทย

- กระทรวงศึกษาธิการ. (2552). *หลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551*. กรุงเทพฯ: โรงพิมพ์ชุมนุมสหกรณ์การเกษตรแห่งประเทศไทย.
- โชติกา ภาษีผล. (2554). *การสร้างและพัฒนาเครื่องมือในการวัดและประเมินผลการศึกษา*. จุฬาลงกรณ์มหาวิทยาลัย.
- โชติกา ภาษีผล, ณีฐภรณ์ หลาวทอง และกมลวรรณ ตังธนกานนท์. (2558). *การวัดและประเมินผล การเรียนรู้*. โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- ณีฐภรณ์ หลาวทอง. (2551). *เอกสารวิชาการ 2757305 การวัดและการประเมินผลการศึกษา*. กรุงเทพมหานคร: คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.
- ณีฐภรณ์ เลขะวัฒนพงษ์. (2560). *การเปรียบเทียบคุณภาพของวิธีการตรวจให้คะแนนความรู้ บางส่วน: การพัฒนาวิธีประยุกต์การให้คะแนนแบบตัดตัวลวง-เลือกตัวถูก*. วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ, จุฬาลงกรณ์มหาวิทยาลัย.
- พัชรี มีวรรณ. (2540). *ผลของจำนวนตัวเลือกที่มีต่อค่าฟังก์ชันสารสนเทศของแบบสอบถามแบบถูกผิดในระดับการศึกษาที่ต่างกัน*. วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ, จุฬาลงกรณ์มหาวิทยาลัย.
- พิชัย ละแมนชัย. (2538). *ขนาดกลุ่มตัวอย่างขั้นต่ำสำหรับการเปรียบเทียบคะแนนในแนวระดับ ตามแนวทฤษฎีการตอบสนองรายข้อระหว่างแบบที่มีแบบแผนการเปรียบเทียบและความยากของแบบสอบถามแตกต่างกัน*. ปริญญาครุศาสตรดุษฎีบัณฑิต, จุฬาลงกรณ์มหาวิทยาลัย.
- ธนบดี อินหาดกรวด และณีฐภรณ์ หลาวทอง. (2560). *การเปรียบเทียบผลการวินิจฉัยนิมโนทัศน์ที่คลาดเคลื่อนในรายวิชาชีววิทยาของนักเรียนชั้นมัธยมศึกษาตอนปลาย*. วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ, จุฬาลงกรณ์มหาวิทยาลัย.
- ธีรุตม์ สุขสกุลวัฒน์. (2563). *การวิเคราะห์โมเดลทฤษฎีการตอบสนองข้อสอบแบบผสมพหุระดับด้วยโมเดลทฤษฎีการตอบสนองข้อสอบทางเลือกที่คำนึงถึงการเดา*. วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ, จุฬาลงกรณ์มหาวิทยาลัย.
- สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี. (2560). *หนังสือเรียนรายวิชาพื้นฐานวิทยาศาสตร์ชีวภาพ ชั้นมัธยมศึกษาปีที่ 4-6*. กรุงเทพฯ: สกสศ. ลาดพร้าว.
- ศิริชัย กาญจนวาสี. (2555). *ทฤษฎีการทดสอบแนวใหม่ (Modern test theories)*. พิมพ์ครั้งที่ 4. กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

- เอมอร จังศิริพรปกรณ์, (2545a). การเปรียบเทียบคุณภาพของแบบสอบเลือกตอบเมื่อตรวจด้วยวิธีการให้คะแนนความรู้บางส่วนกับวิธีประเพณีนิยม. จุฬาลงกรณ์มหาวิทยาลัย.
- เอมอร จังศิริพรปกรณ์, (2545b). การให้คะแนนความรู้บางส่วนทำได้อย่างไร. วารสารครุศาสตร์, 23, 21-32.

ภาษาอังกฤษ

- Baldiga, K. (2014). Gender differences in willingness to guess. *Management Science*, 60(2), 434-448.
- Ben-Simon, A., Budescu, D., & Nevo, B. (1997). A comparative study of measures of partial knowledge in multiple-choice tests. *Applied Psychological Measurement*, 21(1), 65-88.
- Bereby-Meyer, Y., Meyer, J., & Flascher, O. (2002). Prospect theory analysis of guessing in multiple choice tests. *Journal of Behavioral Decision Making*, 15(4), 313-327.
- Bo, Y., Lewis, C., & Budescu, D. (2015). An option-based partial credit item response model. *Quantitative Psychology Research*, 89, 45-72.
- Budescu, D., & Bo, Y. (2015). Analyzing test-taking behavior: decision theory meets psychometric theory. *Psychometrika*, 80(4), 1105-1122.
- Cao, J., & Stokes, S. (2008). Bayesian IRT Guessing Models for Partial Guessing Behaviors. *Psychometrika*, 73(2), 209-230.
- Chang, S-H., Lin, P. C., & Lin, Z. C. (2007). Measures of Partial Knowledge and Unexpected Response in Multiple-Choice Tests. *Educational Technology & Society*, 10(4), 95-109.
- Croson, R., & Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, 47(2), 448-474.
- De Ayala, R. J., (1994). The Influence of Multidimensionality on the Graded Response Model. *Applied Psychological Measurement*, 18, 155-170.
- Hambleton, R., Swaminathan, H., & Rogers, H. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.
- Han, K. (2012). Fixing the c Parameter in the Three-Parameter Logistic Model. *Practical Assessment, Research & Evaluation*, 17(1), 1-24.

- Hutchinson, T. P. (1991). *Ability, partial information and guessing: Statistical modelling applied to multiple-choice tests*. Rundle Mall, South Australia: Rumsby Scientific Publishing.
- Lau, S. H., Hong, K. S., Lau, P. N. K., & Usop, H. (2014). Robustness of Number Right Elimination Testing (NRET) Scoring Method for Multiple-Choice Items in Computer Adaptive Assessment System (CAAS). *Research and Practice in Technology Enhanced Learning*, 9(2), 283-300.
- Liang, T., & Wells, C. S. (2009). A Model Fit Statistic for Generalized Partial Credit Model. *Educational and Psychological Measurement*, 69, 913-928.
- Lord, F. M. (1974). Estimation of latent ability and item parameters when there are omitted responses. *Psychometrika*, 39(2), 247-264.
- Muraki, E. (1992). A Generalized Partial Credit Model: Application of an EM Algorithm. *Applied Psychological Measurement*, 16, 159-176.
- Naumanko, O. (2014). *Comparison of Various Polytomous Item Response Theory Modeling Approaches for Task Based Simulation CPA Exam Data*. The University of North Carolina at Greensboro.
- Reckase, M.D. (2009). *Multidimensional item response theory*. New York, NY: Springer.
- San Martin, E., Del Pino, G., & De Boeck, P. (2006). IRT models for ability-based guessing. *Applied Psychological Measurement*, 30(3), 183-203.
- Smith, R. M. (1993). Guessing and the Rasch model. *Rasch Measurement Transactions*, 6(4), 262-263.
- Tsai, T., & Lin, C. (2015). Modeling Guessing Properties of Multiple-Choice Items in the Measurement of Political Knowledge. *Paper was presented at the 2015 Asian Political Methodology Conference, Jan 9-10, 2015*. Taipei, Taiwan.
- Waugh, K. C. (2013). *Assessment of Student Achievement*. United State of America: Pearson Education.
- Wise, S., & Kong, X. (2005). Response Time Effort: A New Measure of Examinee Motivation in Computer-Based Tests. *Applied Measurement in Education*, 18(2), 163-183.



ภาคผนวก



ผู้ทรงคุณวุฒิด้านการวัดและประเมินผลทางการศึกษา

รองศาสตราจารย์ ดร.กฤษฎากาญจน์ โตพิทักษ์

อาจารย์ประจำภาควิชาบริหาร วิจัย และพัฒนานวัตกรรมทางการศึกษา

คณะศึกษาศาสตร์ มหาวิทยาลัยนเรศวร

ผู้ทรงคุณวุฒิด้านการวิจัย

ผู้ช่วยศาสตราจารย์ ดร.ชำนาญ ปานวงษ์

อาจารย์ประจำภาควิชาบริหาร วิจัย และพัฒนานวัตกรรมทางการศึกษา

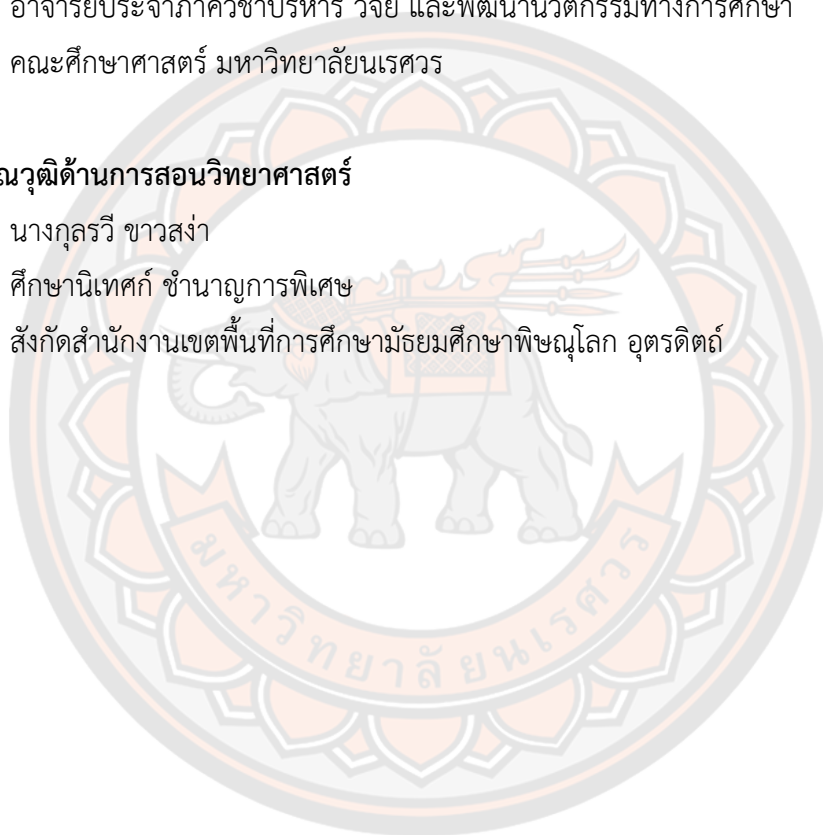
คณะศึกษาศาสตร์ มหาวิทยาลัยนเรศวร

ผู้ทรงคุณวุฒิด้านการสอนวิทยาศาสตร์

นางกุลรวิ ขาวสง่า

ศึกษานิเทศก์ ชำนาญการพิเศษ

สังกัดสำนักงานเขตพื้นที่การศึกษามัธยมศึกษาพิษณุโลก อุตรดิตถ์





ภาคผนวก ข

ผลการตรวจสอบคุณภาพของแบบสอบถามเลือกตอบด้านสมรรถนะทางวิทยาศาสตร์

1. ผลการตรวจสอบความตรงเชิงเนื้อหา (Content Validity) ของแบบสอบถามตอบด้าน
สมรรถนะทางวิทยาศาสตร์ จำนวน 40 ข้อ โดยผู้เชี่ยวชาญจำนวน 3 ท่าน

ข้อ	คนที่ 1	คนที่ 2	คนที่ 3	IOC	ข้อ	คนที่ 1	คนที่ 2	คนที่ 3	IOC
1	1.00	1.00	1.00	1.00	21	1.00	1.00	1.00	1.00
2	0	1.00	1.00	0.67	22	1.00	1.00	1.00	1.00
3	1.00	0	1.00	0.67	23	1.00	1.00	1.00	1.00
4	1.00	1.00	1.00	1.00	24	1.00	0	1.00	0.67
5	1.00	1.00	1.00	1.00	25	1.00	1.00	1.00	1.00
6	1.00	1.00	1.00	1.00	26	1.00	1.00	1.00	1.00
7	1.00	0	1.00	0.67	27	0	1.00	1.00	0.67
8	1.00	1.00	1.00	1.00	28	1.00	1.00	0	0.67
9	1.00	1.00	1.00	1.00	29	1.00	1.00	1.00	1.00
10	1.00	1.00	0	0.67	30	1.00	1.00	1.00	1.00
11	1.00	0	1.00	0.67	31	1.00	1.00	1.00	1.00
12	1.00	1.00	1.00	1.00	32	1.00	1.00	1.00	1.00
13	1.00	1.00	1.00	1.00	33	1.00	1.00	1.00	1.00
14	1.00	1.00	1.00	1.00	34	1.00	1.00	0	0.67
15	1.00	1.00	1.00	1.00	35	1.00	1.00	1.00	1.00
16	1.00	1.00	0	0.67	36	0	1.00	1.00	0.67
17	0	1.00	1.00	0.67	37	1.00	1.00	1.00	1.00
18	0	1.00	1.00	0.67	38	1.00	1.00	1.00	1.00
19	1.00	1.00	1.	1.00	39	1.00	0	1.00	0.67
20	1.00	0	1.0000	0.67	40	1.00	0	1.00	0.67

2. ตัวอย่างผลการวิเคราะห์ความตรงเชิงโครงสร้างของแบบสอบถามสองระดับด้าน สมรรถนะทางวิทยาศาสตร์ ที่มีระดับความมั่นใจ 2 ระดับ ด้วยโปรแกรมสำเร็จรูป Mplus

Mplus VERSION 7.11

MUTHEN & MUTHEN

04/29/2024 9:08 AM

INPUT INSTRUCTIONS

TITLE: CFA1

DATA: FILE IS "C:\Users\Lenovo\Desktop\CFA\DATA_Con1.txt";

VARIABLE: NAMES ARE X1-X20;

USEVARIABLES ARE X1-X20;

ANALYSIS:

TYPE IS GENERAL;

ESTIMATOR IS ML;

ITERATIONS = 1000;

CONVERGENCE = 0.00005;

OUTPUT: SAMPSTAT MODINDICES(3.84) STANDARDIZED;

Model:

exp by X1-X8;

eva by X9-X12;

int by X13-X20;

sc by exp;

sc by eva;

sc by int;

SUMMARY OF ANALYSIS

Number of groups	1
Number of observations	375
Number of dependent variables	20
Number of independent variables	0
Number of continuous latent variables	4

Observed dependent variables

Continuous

X1	X2	X3	X4	X5	X6
X7	X8	X9	X10	X11	X12
X13	X14	X15	X16	X17	X18
X19	X20				

Continuous latent variables

EXP	EVA	INT	SC
-----	-----	-----	----

Estimator	ML
Information matrix	OBSERVED
Maximum number of iterations	1000
Convergence criterion	0.500D-04
Maximum number of steepest descent iterations	20
Input data file(s)	C:\Users\Lenovo\Desktop\CFA\DATA_Con1.txt
Input data format	FREE
MODEL FIT INFORMATION	
Number of Free Parameters	122
Loglikelihood	
H0 Value	-8255.525
H1 Value	-8189.823
Information Criteria	
Akaike (AIC)	16755.051
Bayesian (BIC)	17234.135
Sample-Size Adjusted BIC	16847.062
(n* = (n + 2) / 24)	
Chi-Square Test of Model Fit	
Value	131.404
Degrees of Freedom	108
P-Value	0.0624

RMSEA (Root Mean Square Error Of Approximation)

Estimate	0.024
90 Percent C.I.	0.000 0.037
Probability RMSEA \leq .05	1.000

CFI/TLI

CFI	0.989
TLI	0.980

Chi-Square Test of Model Fit for the Baseline Model

Value	2273.474
Degrees of Freedom	190
P-Value	0.0000

SRMR (Standardized Root Mean Square Residual)

Value	0.034
-------	-------

STANDARDIZED MODEL RESULTS

STDYX Standardization

		Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
EXP	BY				
	X1	0.488	0.042	11.559	0.000
	X2	0.544	0.040	13.508	0.000
	X3	0.617	0.037	16.641	0.000
	X4	0.332	0.052	6.375	0.000
	X5	0.394	0.044	8.963	0.000
	X6	0.590	0.036	16.562	0.000
	X7	0.461	0.043	10.674	0.000
	X8	0.594	0.037	15.958	0.000
EVA	BY				
	X9	0.567	0.042	13.420	0.000
	X10	0.389	0.050	7.761	0.000
	X11	0.743	0.041	18.300	0.000
	X12	0.176	0.055	3.216	0.001

INT	BY				
	X13	0.575	0.039	14.677	0.000
	X14	0.446	0.048	9.310	0.000
	X15	0.467	0.042	11.067	0.000
	X16	0.458	0.041	11.229	0.000
	X17	0.463	0.045	10.390	0.000
	X18	0.155	0.051	3.022	0.003
	X19	0.573	0.042	13.546	0.000
	X20	0.084	0.052	1.611	0.107
SC	BY				
	EXP	1.051	0.032	33.341	0.000
	EVA	0.861	0.043	20.151	0.000
	INT	1.075	0.041	26.250	0.000
Intercepts					
	X1	0.768	0.059	13.096	0.000
	X2	0.738	0.058	12.682	0.000
	X3	0.749	0.058	12.837	0.000
	X4	0.997	0.063	15.912	0.000
	X5	0.616	0.056	10.933	0.000
	X6	0.666	0.057	11.659	0.000
	X7	1.252	0.069	18.122	0.000
	X8	0.531	0.055	9.626	0.000
	X9	0.770	0.059	13.110	0.000
	X10	0.988	0.063	15.711	0.000
	X11	0.799	0.059	13.484	0.000
	X12	0.481	0.054	8.836	0.000
	X13	0.843	0.060	14.053	0.000
	X14	1.031	0.063	16.373	0.000
	X15	0.619	0.056	11.009	0.000
	X16	0.811	0.060	13.633	0.000
	X17	0.943	0.062	15.237	0.000

X18	0.739	0.058	12.717	0.000
X19	0.649	0.057	11.427	0.000
X20	0.679	0.057	11.848	0.000
Variances				
SC	1.000	0.000	999.000	999.000
Residual Variances				
X1	0.762	0.041	18.468	0.000
X2	0.704	0.044	16.031	0.000
X3	0.619	0.046	13.505	0.000
X4	0.890	0.035	25.794	0.000
X5	0.845	0.035	24.369	0.000
X6	0.651	0.042	15.469	0.000
X7	0.788	0.040	19.794	0.000
X8	0.647	0.044	14.626	0.000
X9	0.678	0.048	14.137	0.000
X10	0.849	0.039	21.811	0.000
X11	0.448	0.060	7.435	0.000
X12	0.969	0.019	50.586	0.000
X13	0.669	0.045	14.822	0.000
X14	0.801	0.043	18.744	0.000
X15	0.782	0.039	19.841	0.000
X16	0.790	0.037	21.158	0.000
X17	0.786	0.041	19.070	0.000
X18	0.976	0.016	61.682	0.000
X19	0.672	0.048	13.873	0.000
X20	0.993	0.009	113.752	0.000
EXP	-0.105	999.000	999.000	999.000
EVA	0.258	0.074	3.506	0.000
INT	-0.155	999.000	999.000	999.000

R-SQUARE

Observed				Two-Tailed
Variable	Estimate	S.E.	Est./S.E.	P-Value
X1	0.238	0.041	5.779	0.000
X2	0.296	0.044	6.754	0.000
X3	0.381	0.046	8.321	0.000
X4	0.110	0.035	3.187	0.001
X5	0.155	0.035	4.482	0.000
X6	0.349	0.042	8.281	0.000
X7	0.212	0.040	5.337	0.000
X8	0.353	0.044	7.979	0.000
X9	0.322	0.048	6.710	0.000
X10	0.151	0.039	3.881	0.000
X11	0.552	0.060	9.150	0.000
X12	0.031	0.019	1.608	0.108
X13	0.331	0.045	7.338	0.000
X14	0.199	0.043	4.655	0.000
X15	0.218	0.039	5.534	0.000
X16	0.210	0.037	5.615	0.000
X17	0.214	0.041	5.195	0.000
X18	0.024	0.016	1.511	0.131
X19	0.328	0.048	6.773	0.000
X20	0.007	0.009	0.806	0.420
Latent				Two-Tailed
Variable	Estimate	S.E.	Est./S.E.	P-Value
EXP	Undefined	0.11045E+01		
EVA	0.742	0.074	10.075	0.000
INT	Undefined	0.11548E+01		

QUALITY OF NUMERICAL RESULTS

Condition Number for the Information Matrix (ratio of smallest to largest eigenvalue)	0.286E-03
--	-----------

Diagram output

c:\users\lenovo\desktop\cfa\output1.dgm

Beginning Time: 09:08:00

Ending Time: 09:08:00

Elapsed Time: 00:00:00



3. ตัวอย่างผลการวิเคราะห์คุณภาพของแบบสอบเลือกตอบด้านสมรรถนะทางวิทยาศาสตร์ ที่มีระดับความมั่นใจ 2 ระดับ โดยใช้ Generalized Partial Credit Model (G-PCM) ด้วยโปรแกรมสำเร็จรูป IRTPRO4 Student

IRTPRO Version 4.1

Output generated by IRTPRO estimation engine Version 5.20 (32-bit)

Project:	Polytomous IRT
Description:	Polytomous Scoring
Date:	29 April 2024
Time:	09:51 AM

Table of Contents

Nominal Model Slopes and Scoring Function Contrasts for Group 1, logit: $(a s_k \psi + c_k)$;

$s = T\alpha$

Nominal Model Scoring Function Values for Group 1, logit: $(a s_k \psi + c_k)$; $s = T\alpha$

Nominal Model Intercept Contrasts for Group 1, logit: $(a s_k \psi + c_k)$; $c = T\gamma$

Original (Bock, 1972) Parameters, Nominal Items for Group 1, logit: $(a_k \psi + c_k)$

Summed-Score Based Item Diagnostic Tables and X^2 s for Group 1

Group Parameter Estimates

Marginal fit (X^2) and Standardized LD X^2 Statistics for Group 1

Item Information Function Values for Group 1 at 15 Values of θ from -2.8 to 2.8

Likelihood-based Values and Goodness of Fit Statistics

Summary of the Data and Control Parameters

GPC Model Item Parameter Estimates, logit: $a[k(\psi - b) + \sum d_k]$

Item	Label	a	$s.e.$	b	$s.e.$	d_1	d_2	$s.e.$	d_3	$s.e.$
1	X1	¹ 0.71	0.10	0.77	0.14	0.00	-0.97	0.29	0.97	0.29
2	X2	⁴ 0.94	0.13	0.67	0.11	0.00	-0.71	0.23	0.71	0.23
3	X3	⁷ 0.83	0.12	0.64	0.12	0.00	-1.20	0.31	1.20	0.31

Item	Label		a	$s.e.$	b	$s.e.$	d_1	d_2	$s.e.$	d_3	$s.e.$
4	X4	¹⁰	0.71	0.11	0.31	0.12	0.00	-0.38	0.21	0.38	0.21
5	X5	¹³	0.50	0.08	1.44	0.24	0.00	-2.66	0.64	2.66	0.64
6	X6	¹⁶	1.02	0.14	0.82	0.11	0.00	-0.65	0.22	0.65	0.22
7	X7	¹⁹	0.65	0.11	-0.35	0.12	0.00	-1.15	0.33	1.15	0.33
8	X8	²²	0.85	0.12	1.24	0.15	0.00	-1.63	0.39	1.63	0.39
9	X9	²⁵	0.89	0.13	0.67	0.12	0.00	-0.59	0.22	0.59	0.22
10	X10	²⁸	0.65	0.11	0.24	0.12	0.00	-0.91	0.29	0.91	0.29
11	X11	³¹	1.19	0.17	0.51	0.09	0.00	-0.26	0.15	0.26	0.15
12	X12	³⁴	0.24	0.08	3.72	1.13	0.00	-7.91	2.77	7.91	2.77
13	X13	³⁷	0.91	0.13	0.46	0.10	0.00	-0.63	0.22	0.63	0.22
14	X14	⁴⁰	0.67	0.11	0.20	0.11	0.00	-0.72	0.26	0.72	0.26
15	X15	⁴³	0.75	0.11	1.24	0.17	0.00	-0.86	0.29	0.86	0.29
16	X16	⁴⁶	0.69	0.10	0.62	0.13	0.00	-1.16	0.32	1.16	0.32
17	X17	⁴⁹	0.84	0.13	0.33	0.10	0.00	-0.41	0.20	0.41	0.20
18	X18	⁵²	0.14	0.07	3.11	1.46	0.00	-8.22	4.02	8.22	4.02
19	X19	⁵⁵	0.65	0.10	1.07	0.16	0.00	-1.83	0.44	1.83	0.44
20	X20	⁵⁸	0.06	0.07	8.75	9.65	0.00	-20.55	22.90	20.55	22.90

Nominal Model Slopes and Scoring Function Contrasts for Group 1, logit: $(a s_k \theta + c_k)$; $s = T\alpha$ (Back to TOC)

Item	Label		a	$s.e.$ Contrasts	α_1	$s.e.$	α_2	$s.e.$
1	X1	¹	0.71	0.10 Trend	1.00	----	0.00	----
2	X2	⁴	0.94	0.13 Trend	1.00	----	0.00	----
3	X3	⁷	0.83	0.12 Trend	1.00	----	0.00	----
4	X4	¹⁰	0.71	0.11 Trend	1.00	----	0.00	----
5	X5	¹³	0.50	0.08 Trend	1.00	----	0.00	----
6	X6	¹⁶	1.02	0.14 Trend	1.00	----	0.00	----
7	X7	¹⁹	0.65	0.11 Trend	1.00	----	0.00	----
8	X8	²²	0.85	0.12 Trend	1.00	----	0.00	----
9	X9	²⁵	0.89	0.13 Trend	1.00	----	0.00	----
10	X10	²⁸	0.65	0.11 Trend	1.00	----	0.00	----

Item	Label	a	s.e. Contrasts	α_1	s.e.	α_2	s.e.
11	X11	³¹ 1.19	0.17 Trend	1.00	----	0.00	----
12	X12	³⁴ 0.24	0.08 Trend	1.00	----	0.00	----
13	X13	³⁷ 0.91	0.13 Trend	1.00	----	0.00	----
14	X14	⁴⁰ 0.67	0.11 Trend	1.00	----	0.00	----
15	X15	⁴³ 0.75	0.11 Trend	1.00	----	0.00	----
16	X16	⁴⁶ 0.69	0.10 Trend	1.00	----	0.00	----
17	X17	⁴⁹ 0.84	0.13 Trend	1.00	----	0.00	----
18	X18	⁵² 0.14	0.07 Trend	1.00	----	0.00	----
19	X19	⁵⁵ 0.65	0.10 Trend	1.00	----	0.00	----
20	X20	⁵⁸ 0.06	0.07 Trend	1.00	----	0.00	----

Nominal Model Scoring Function Values for Group 1, logit: $(a s_k \theta + c_k)$; $s =$

T α ([Back to TOC](#))

Item	Category	s_1	s_2	s_3
1	X1	0.00	1.00	2.00
2	X2	0.00	1.00	2.00
3	X3	0.00	1.00	2.00
4	X4	0.00	1.00	2.00
5	X5	0.00	1.00	2.00
6	X6	0.00	1.00	2.00
7	X7	0.00	1.00	2.00
8	X8	0.00	1.00	2.00
9	X9	0.00	1.00	2.00
10	X10	0.00	1.00	2.00
11	X11	0.00	1.00	2.00
12	X12	0.00	1.00	2.00
13	X13	0.00	1.00	2.00
14	X14	0.00	1.00	2.00
15	X15	0.00	1.00	2.00
16	X16	0.00	1.00	2.00
17	X17	0.00	1.00	2.00

Item	Category	s_1	s_2	s_3
18	X18	0.00	1.00	2.00
19	X19	0.00	1.00	2.00
20	X20	0.00	1.00	2.00

Nominal Model Intercept Contrasts for Group 1, logit: $(a s_k \boldsymbol{\psi} + c_k)$; $c = T\boldsymbol{\gamma}$ ([Back to TOC](#))

Item	Label	Contrasts	$\boldsymbol{\psi}_1$	<i>s.e.</i>	$\boldsymbol{\psi}_2$	<i>s.e.</i>
1	X1	Trend	² -0.55	0.08	³ -0.69	0.15
2	X2	Trend	⁵ -0.64	0.10	⁶ -0.67	0.16
3	X3	Trend	⁸ -0.53	0.09	⁹ -0.99	0.17
4	X4	Trend	¹¹ -0.22	0.08	¹² -0.27	0.13
5	X5	Trend	¹⁴ -0.71	0.08	¹⁵ -1.32	0.19
6	X6	Trend	¹⁷ -0.84	0.11	¹⁸ -0.67	0.17
7	X7	Trend	²⁰ 0.23	0.08	²¹ -0.75	0.14
8	X8	Trend	²³ -1.05	0.11	²⁴ -1.38	0.22
9	X9	Trend	²⁶ -0.59	0.09	²⁷ -0.53	0.15
10	X10	Trend	²⁹ -0.15	0.07	³⁰ -0.59	0.14
11	X11	Trend	³² -0.61	0.11	³³ -0.31	0.16
12	X12	Trend	³⁵ -0.89	0.08	³⁶ -1.88	0.24
13	X13	Trend	³⁸ -0.42	0.09	³⁹ -0.57	0.15
14	X14	Trend	⁴¹ -0.13	0.07	⁴² -0.48	0.14
15	X15	Trend	⁴⁴ -0.93	0.10	⁴⁵ -0.65	0.17
16	X16	Trend	⁴⁷ -0.43	0.08	⁴⁸ -0.80	0.15
17	X17	Trend	⁵⁰ -0.28	0.08	⁵¹ -0.34	0.14
18	X18	Trend	⁵³ -0.44	0.06	⁵⁴ -1.17	0.16
19	X19	Trend	⁵⁶ -0.69	0.09	⁵⁷ -1.19	0.18
20	X20	Trend	⁵⁹ -0.53	0.06	⁶⁰ -1.24	0.17

Original (Bock, 1972) Parameters, Nominal Items for Group 1, logit: $(\alpha_k \theta + c_k)$ [\(Back to TOC\)](#)

Item	Label	Category	1	2	3
1	X1	<i>a</i>	0.00	0.71	1.41
		<i>c</i>	-0.00	-1.24	-1.09
2	X2	<i>a</i>	0.00	0.94	1.89
		<i>c</i>	-0.00	-1.31	-1.27
3	X3	<i>a</i>	0.00	0.83	1.65
		<i>c</i>	-0.00	-1.52	-1.05
4	X4	<i>a</i>	0.00	0.71	1.42
		<i>c</i>	-0.00	-0.49	-0.44
5	X5	<i>a</i>	0.00	0.50	0.99
		<i>c</i>	-0.00	-2.03	-1.42
6	X6	<i>a</i>	0.00	1.02	2.05
		<i>c</i>	-0.00	-1.51	-1.69
7	X7	<i>a</i>	0.00	0.65	1.30
		<i>c</i>	0.00	-0.52	0.46
8	X8	<i>a</i>	0.00	0.85	1.69
		<i>c</i>	-0.00	-2.43	-2.10
9	X9	<i>a</i>	0.00	0.89	1.77
		<i>c</i>	-0.00	-1.12	-1.18
10	X10	<i>a</i>	0.00	0.65	1.30
		<i>c</i>	-0.00	-0.75	-0.31
11	X11	<i>a</i>	0.00	1.19	2.39
		<i>c</i>	-0.00	-0.92	-1.22
12	X12	<i>a</i>	0.00	0.24	0.48
		<i>c</i>	-0.00	-2.77	-1.77
13	X13	<i>a</i>	0.00	0.91	1.83
		<i>c</i>	-0.00	-0.99	-0.84
14	X14	<i>a</i>	0.00	0.67	1.35
		<i>c</i>	-0.00	-0.62	-0.26
15	X15	<i>a</i>	0.00	0.75	1.50

Item	Label	Category	1	2	3
		<i>c</i>	-0.00	-1.58	-1.86
16	X16	<i>a</i>	0.00	0.69	1.37
		<i>c</i>	-0.00	-1.23	-0.85
17	X17	<i>a</i>	0.00	0.84	1.67
		<i>c</i>	-0.00	-0.62	-0.56
18	X18	<i>a</i>	0.00	0.14	0.29
		<i>c</i>	-0.00	-1.61	-0.89
19	X19	<i>a</i>	0.00	0.65	1.30
		<i>c</i>	-0.00	-1.88	-1.38
20	X20	<i>a</i>	0.00	0.06	0.12
		<i>c</i>	-0.00	-1.77	-1.06

Summed-Score Based Item Diagnostic Tables and χ^2 s for Group 1 ([Back to TOC](#))

S- χ^2 Item Level Diagnostic Statistics

Item	Label	χ^2	<i>d.f.</i>	Probability
1	X1	73.51	40	0.0010
2	X2	96.24	37	0.0001
3	X3	60.68	38	0.0111
4	X4	80.16	41	0.0002
5	X5	51.58	41	0.1242
6	X6	71.08	33	0.0001
7	X7	67.16	36	0.0012
8	X8	53.50	32	0.0100
9	X9	94.81	39	0.0001
10	X10	94.90	38	0.0001
11	X11	67.97	37	0.0014
12	X12	66.18	33	0.0005
13	X13	73.21	36	0.0002
14	X14	84.86	37	0.0001

Item	Label	χ^2	df.	Probability
15	X15	66.19	36	0.0016
16	X16	65.40	42	0.0118
17	X17	120.87	37	0.0001
18	X18	73.12	42	0.0021
19	X19	68.19	38	0.0019
20	X20	64.70	40	0.0080

Group Parameter Estimates [\(Back to TOC\)](#)

Group	Label	μ	s.e.	σ^2	s.e.	σ	s.e.
1	Group 1	0.00	----	1.00	----	1.00	----

Marginal fit (χ^2) and Standardized LD χ^2 Statistics for Group 1 [\(Back to TOC\)](#)

Item	Label	Marginal χ^2	Standardized LD χ^2 Statistics											
			1	2	3	4	5	6	7	8	9	10		
1	X1	0.2												
2	X2	0.7	4.0											
3	X3	0.5	5.8	9.8										
4	X4	0.2	5.4	4.4	6.2									
5	X5	0.0	1.8	2.7	1.9	2.6								
6	X6	0.7	6.0	4.3	4.1	3.4	0.8							
7	X7	0.1	5.7	0.8	1.1	6.1	0.3	2.7						
8	X8	0.2	3.0	3.8	8.0	5.3	3.2	4.9	1.9					
9	X9	0.5	6.6	9.6	1.1	11.5	3.6	12.7	1.8	2.7				
10	X10	0.2	10.2	2.3	5.3	9.2	3.6	1.8	5.2	3.2	14.3			
11	X11	1.2	6.5	5.1	6.1	5.5	2.2	1.4	3.3	3.4	3.5	3.6		
12	X12	0.0	4.5	2.4	5.3	7.7	4.2	3.2	0.9	2.7	3.5	9.6		
13	X13	0.6	4.5	6.7	7.0	10.9	2.9	4.4	0.8	3.7	7.1	7.7		
14	X14	0.2	4.4	12.0	20.9	3.8	3.6	3.3	1.3	7.1	2.6	4.2		
15	X15	0.1	5.7	2.6	7.4	1.8	1.7	2.1	4.1	1.4	9.5	3.0		
16	X16	0.3	4.3	3.0	6.5	3.9	1.0	1.6	-0.7	3.6	2.8	3.8		
17	X17	0.4	5.2	3.9	4.7	11.7	-0.1	4.6	8.9	4.6	15.8	1.4		
18	X18	0.0	1.6	2.1	7.0	1.0	12.2	1.0	-0.8	-0.4	1.5	8.9		

19	X19	0.1	4.0	3.1	5.1	10.8	5.5	7.1	-0.1	6.8	7.7	7.9
20	X20	0.0	2.0	3.3	3.1	0.8	-1.0	0.9	-0.9	2.3	1.7	2.6

Item	Label	Marginal										
		χ^2	11	12	13	14	15	16	17	18	19	
11	X11	1.2										
12	X12	0.0	3.5									
13	X13	0.6	7.7	0.8								
14	X14	0.2	17.6	4.3	11.2							
15	X15	0.1	2.9	4.8	1.6	2.8						
16	X16	0.3	1.1	1.0	5.3	6.4	9.7					
17	X17	0.4	1.7	3.7	3.1	3.0	12.2	3.9				
18	X18	0.0	4.9	16.5	2.2	7.2	-0.0	0.1	2.4			
19	X19	0.1	6.1	6.6	4.7	2.5	4.3	1.7	3.5	-0.2		
20	X20	0.0	-0.6	0.6	0.8	0.3	0.3	0.3	1.3	1.6	3.4	

Item Information Function Values for Group 1 at 15 Values of θ from -2.8 to 2.8 [\(Back to TOC\)](#)

		θ :														
Item	Label	-2.8	-2.4	-2.0	-1.6	-1.2	-0.8	-0.4	-0.0	0.4	0.8	1.2	1.6	2.0	2.4	2.8
1	X1	0.03	0.04	0.06	0.09	0.13	0.19	0.26	0.33	0.38	0.40	0.37	0.32	0.25	0.18	0.13
2	X2	0.02	0.03	0.05	0.09	0.15	0.24	0.38	0.54	0.68	0.70	0.60	0.44	0.29	0.18	0.11
3	X3	0.02	0.04	0.06	0.09	0.15	0.23	0.35	0.47	0.56	0.57	0.49	0.37	0.25	0.16	0.10
4	X4	0.06	0.08	0.11	0.15	0.20	0.26	0.32	0.35	0.36	0.34	0.29	0.23	0.18	0.13	0.09
5	X5	0.02	0.03	0.04	0.05	0.07	0.09	0.12	0.15	0.18	0.20	0.21	0.22	0.20	0.18	0.15
6	X6	0.02	0.02	0.04	0.07	0.12	0.20	0.34	0.53	0.74	0.84	0.76	0.56	0.36	0.21	0.12
7	X7	0.09	0.12	0.17	0.22	0.28	0.32	0.34	0.33	0.29	0.24	0.18	0.13	0.09	0.07	0.05
8	X8	0.01	0.01	0.02	0.04	0.06	0.11	0.18	0.29	0.43	0.57	0.64	0.59	0.46	0.32	0.20
9	X9	0.03	0.04	0.06	0.10	0.16	0.24	0.36	0.49	0.58	0.60	0.52	0.40	0.28	0.18	0.12
10	X10	0.05	0.08	0.11	0.15	0.19	0.25	0.29	0.32	0.33	0.30	0.26	0.20	0.15	0.11	0.08
11	X11	0.02	0.04	0.06	0.11	0.19	0.33	0.56	0.84	1.03	0.97	0.72	0.45	0.26	0.15	0.08
12	X12	0.01	0.01	0.01	0.02	0.02	0.02	0.02	0.03	0.03	0.04	0.04	0.04	0.05	0.05	0.05

θ :		-2.8	-2.4	-2.0	-1.6	-1.2	-0.8	-0.4	-0.0	0.4	0.8	1.2	1.6	2.0	2.4	2.8
13	X13	0.03	0.05	0.08	0.12	0.20	0.30	0.44	0.58	0.65	0.61	0.49	0.34	0.22	0.14	0.09
14	X14	0.06	0.08	0.11	0.16	0.21	0.26	0.31	0.34	0.34	0.31	0.26	0.21	0.15	0.11	0.08
15	X15	0.02	0.03	0.04	0.06	0.09	0.13	0.19	0.26	0.34	0.41	0.45	0.42	0.36	0.28	0.20
16	X16	0.03	0.05	0.07	0.10	0.15	0.21	0.28	0.34	0.38	0.38	0.34	0.28	0.22	0.16	0.11
17	X17	0.05	0.07	0.10	0.15	0.22	0.32	0.42	0.49	0.52	0.47	0.38	0.28	0.20	0.13	0.09
18	X18	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
19	X19	0.02	0.03	0.04	0.06	0.10	0.14	0.19	0.25	0.32	0.36	0.36	0.33	0.28	0.21	0.15
20	X20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Test	1.60	1.86	2.26	2.85	3.69	4.86	6.35	7.96	9.15	9.32	8.39	6.84	5.27	3.97	3.02
Information	1.60	1.86	2.26	2.85	3.69	4.86	6.35	7.96	9.15	9.32	8.39	6.84	5.27	3.97	3.02
:															
Expected	0.79	0.73	0.66	0.59	0.52	0.45	0.40	0.35	0.33	0.33	0.35	0.38	0.44	0.50	0.58
s.e.:															

Marginal Reliability for Response Pattern Scores: 0.83

Likelihood-based Values and Goodness of Fit Statistics ([Back to TOC](#))

Statistics based on the loglikelihood	
-2loglikelihood:	12496.61
Akaike Information Criterion (AIC):	12616.61
Bayesian Information Criterion (BIC):	12852.22

Statistics based on the full item x item x ... classification

The table is too sparse to compute the general multinomial goodness of fit statistics.

Statistics based on one- and two-way marginal tables

M_2 statistics not available for this estimation method.

Summary of the Data and Control Parameters [\(Back to TOC\)](#)

Sample Size	375
Number of Items	20
Number of Dimensions	1

Item	Label	Categories	Model	T_a	T_c
1	X1	3	Nominal	Trend	Trend
2	X2	3	Nominal	Trend	Trend
3	X3	3	Nominal	Trend	Trend
4	X4	3	Nominal	Trend	Trend
5	X5	3	Nominal	Trend	Trend
6	X6	3	Nominal	Trend	Trend
7	X7	3	Nominal	Trend	Trend
8	X8	3	Nominal	Trend	Trend
9	X9	3	Nominal	Trend	Trend
10	X10	3	Nominal	Trend	Trend
11	X11	3	Nominal	Trend	Trend
12	X12	3	Nominal	Trend	Trend
13	X13	3	Nominal	Trend	Trend
14	X14	3	Nominal	Trend	Trend
15	X15	3	Nominal	Trend	Trend
16	X16	3	Nominal	Trend	Trend
17	X17	3	Nominal	Trend	Trend
18	X18	3	Nominal	Trend	Trend
19	X19	3	Nominal	Trend	Trend
20	X20	3	Nominal	Trend	Trend

Parameter Estimation Control Values

Bock-Aitkin EM Algorithm	
Maximum number of cycles:	500
Convergence criterion:	1.00e-005
Maximum number of M-step iterations:	50

Convergence criterion for iterative M-steps:	1.00e-006	
Number of rectangular quadrature points:	49	
Minimum, Maximum quadrature points:	-6.00	6.00
SEM algorithm tolerance:	1.00e-003	
Standard error computation algorithm:	Supplemented EM	

Miscellaneous Control Values

Print parameter numbers?	Yes
Z tolerance, max. abs. logit value:	50.00
Number of processor cores used:	1
Number of cycles completed:	54
Maximum parameter change:	0.00e+000
Number of free parameters:	60

Processing times (in seconds)

E-step computations:	0.02
M-step computations:	0.04
Standard error computations:	0.47
Goodness-of-fit statistics:	0.05
Total:	0.57

Output Files

HTML results and control parameters: C:\Users\Lenovo\Desktop\VRT\Output DATA1.CON1-irt.htm

Convergence and Numerical Stability

Engine status:	Normal termination
SEM algorithm status:	Normal
First-order test:	Convergence criteria satisfied
Condition number of information matrix:	1.70e+001
Second-order test:	Solution is a possible local maximum



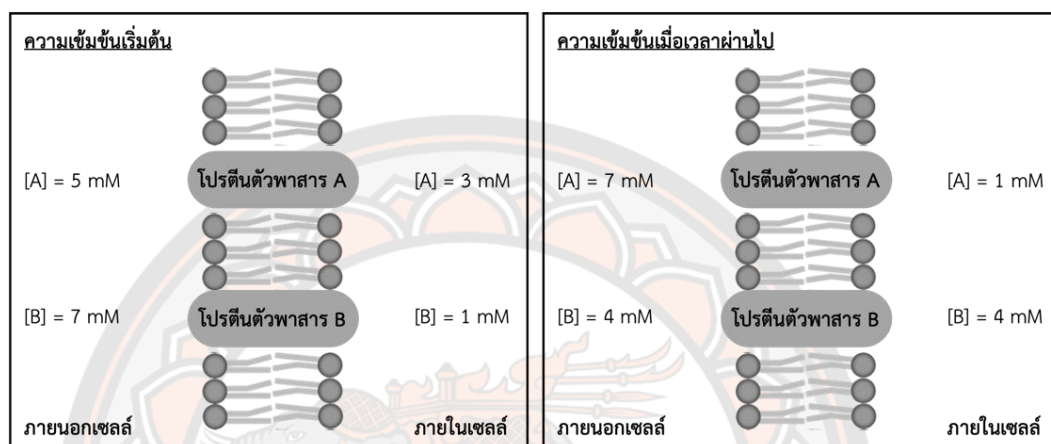
ภาคผนวก ค

ตัวอย่างแบบสอบถามเลือกตอบด้านสมรรถนะทางวิทยาศาสตร์
ที่มีระดับความมั่นใจที่แตกต่างกัน ในรูปแบบออนไลน์

ตัวอย่างแบบสอบเลือกตอบด้านสมรรถนะทางวิทยาศาสตร์ที่มีระดับความมั่นใจ 2 ระดับ

องค์ประกอบที่ 1 การอธิบายปรากฏการณ์ในเชิงวิทยาศาสตร์

1. ในการลำเลียงสาร A และสาร B โดยอาศัยโปรตีนตัวพา เป็นดังภาพ



จากข้อมูลข้างต้น การลำเลียงสาร A และสาร B อาศัยกลไกการลำเลียงสารแบบใด ตามลำดับ

- active transport และ simple diffusion
- simple diffusion และ facilitated diffusion
- facilitated diffusion และ active transport
- active transport และ facilitated diffusion

กระดาษคำตอบ

ข้อ	ตัวเลือก				ระดับความมั่นใจ	
	ก	ข	ค	ง	มั่นใจ	ไม่มั่นใจ
1						

2. ข้อใดกล่าวถึงการรักษาคุณภาพของกรด-เบสในเลือดได้ถูกต้อง
- ถ้าเลือดเป็นเบส อัตราการหายใจจะลดลง เพื่อลดปริมาณ H^+
 - ถ้าเลือดเป็นเบส อัตราการหายใจจะเพิ่มขึ้น เพื่อเพิ่มปริมาณ H^+
 - ถ้าเลือดเป็นกรด อัตราการหายใจจะลดลง เพื่อเพิ่มปริมาณ CO_2
 - ถ้าเลือดเป็นกรด อัตราการหายใจจะเพิ่มขึ้น เพื่อลดปริมาณ CO_2

กระดาษคำตอบ

ข้อ	ตัวเลือก				ระดับความมั่นใจ	
	ก	ข	ค	ง	มั่นใจ	ไม่มั่นใจ
2						

องค์ประกอบที่ 2 การประเมินและออกแบบกระบวนการสืบเสาะหาความรู้ทางวิทยาศาสตร์

3. เมื่อบดเนื้อเยื่อของพืชหัวชนิดหนึ่งกับน้ำจะได้ของเหลวขาวขุ่นคล้ายน้ำแป้ง ถ้าต้องการทดสอบว่าของเหลวนั้นมีแป้งเป็นองค์ประกอบหรือไม่ ควรทำการทดลองโดยใช้หลอดไตบ่าง จากตารางการทดลองข้างล่างนี้

หลอดที่	ของเหลวจากพืชหัว	น้ำ	น้ำแป้งมัน	นมจืด	สารละลายไอโอดีน	สารละลายไบยูเรต
1	+	-	-	-	+	-
2	+	-	-	-	-	+
3	-	-	+	-	+	-
4	-	-	-	+	-	+
5	-	+	-	-	+	-
6	-	+	-	-	-	+

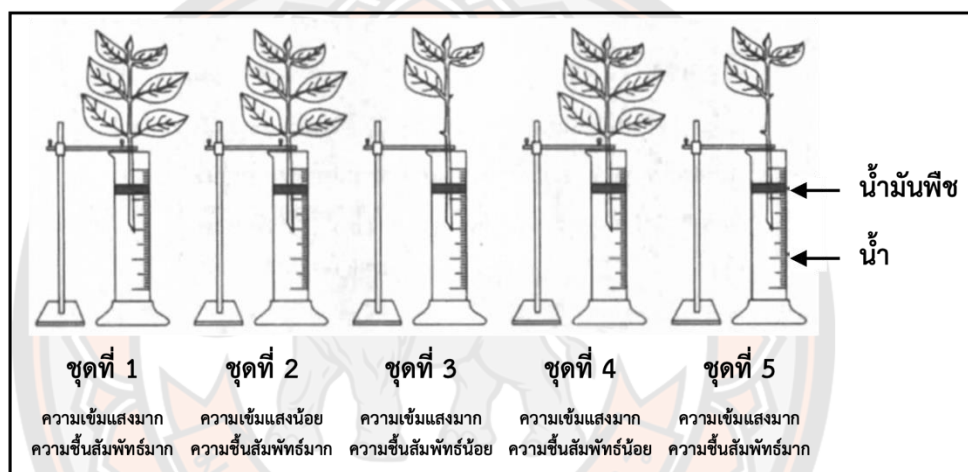
กำหนด: + มีการใช้สาร - ไม่มีการใช้สาร

- หลอดที่ 1 2
- หลอดที่ 2 6
- หลอดที่ 1 3 5
- หลอดที่ 2 4 6

กระดาษคำตอบ

ข้อ	ตัวเลือก				ระดับความมั่นใจ	
	ก	ข	ค	ง	มั่นใจ	ไม่มั่นใจ
3						

4. นักเรียนจัดชุดการทดลอง 5 ชุด โดยใช้กิ่งไม้ที่มีอายุเท่ากันจากต้นเดียวกัน ตัดใบในชุดการทดลองที่ 3 และ 5 ออกบางส่วน จากนั้นแช่กิ่งไม้ในหลอดทดลองที่มีน้ำ 32 มิลลิลิตร และมีน้ำมันพืช 3 มิลลิลิตร เททับอยู่ แล้วตั้งทิ้งไว้ในสภาวะที่แตกต่างกัน ดังภาพ



เมื่อเวลาผ่านไป 30 นาที พบว่า ปริมาณน้ำคงเหลือในหลอดทดลอง ดังตาราง

ชุดการทดลองที่	ปริมาณน้ำคงเหลือในหลอดทดลอง (ml)
1	15
2	20
3	25
4	10
5	30

ข้อใดเลือกชุดการทดลองเพื่อศึกษาปัจจัยที่เกี่ยวข้องกับการคายน้ำและเปรียบเทียบอัตราการคายน้ำได้ถูกต้อง

	ชุดการทดลองที่ใช้	ปัจจัยที่ต้องการศึกษา	ผลการเปรียบเทียบอัตราการคายน้ำ
ก.	1 และ 2	ความเข้มแสง	ชุดการทดลองที่ 2 มีอัตราการคายน้ำมากกว่า ชุดการทดลองที่ 1
ข.	1 และ 4	ความชื้นสัมพัทธ์	ชุดการทดลองที่ 4 มีอัตราการคายน้ำมากกว่า ชุดการทดลองที่ 1
ค.	3 และ 4	จำนวนใบ	ชุดการทดลองที่ 3 มีอัตราการคายน้ำมากกว่า ชุดการทดลองที่ 4
ง.	3 และ 5	ความเข้มแสง	ชุดการทดลองที่ 5 มีอัตราการคายน้ำมากกว่า ชุดการทดลองที่ 3

กระดาษคำตอบ

ข้อ	ตัวเลือก				ระดับความมั่นใจ	
	ก	ข	ค	ง	มั่นใจ	ไม่มั่นใจ
4						

องค์ประกอบที่ 3 การแปลความหมายข้อมูลและการใช้ประจักษ์พยานในเชิงวิทยาศาสตร์

5. โรคเบาหวานชนิดที่ 1 เกิดจากตับอ่อนสร้าง “ฮอร์โมนอินซูลิน” ได้น้อย หรือสร้างไม่ได้เลย ซึ่งฮอร์โมนชนิดนี้ทำหน้าที่ช่วยให้ร่างกายสลายน้ำตาลมาใช้เป็นพลังงาน เมื่ออินซูลินในร่างกายไม่เพียงพอ ร่างกายไม่สามารถนำน้ำตาลไปใช้ได้ ทำให้ปริมาณน้ำตาลในเลือดสูง ร่างกายจะปรับตัวไปใช้พลังงานจากการสลายสารอาหารอื่น เช่น ไขมัน หรือ โปรตีน ซึ่งก่อให้เกิดการสะสมของของเสียที่เป็นอันตรายในเลือด น้ำตาลและของเสียปริมาณมากทำให้เลือดของผู้ป่วยโรคเบาหวานมีความเข้มข้นสูง ร่างกายจึงต้องขับน้ำตาลและของเสียผ่านทางปัสสาวะ ทำให้ผู้ป่วยโรคเบาหวานปัสสาวะบ่อยในปริมาณมากและรู้สึกกระหายน้ำ

ข้อใดไม่ใช่สาเหตุที่ทำให้ผู้ป่วยโรคเบาหวานปัสสาวะบ่อยในปริมาณมาก

- ก. น้ำจะออสโมซิสจากเซลล์มาสู่เลือด
ข. ร่างกายกำจัดน้ำตาลที่มีมากเกินไปในเลือด

- ค. ต่อมได้สมองหลังฮอโมนเข้าสู่กระแสเลือดไปกระตุ้นต่อหน่วยไต
ง. ของเหลวที่ผ่านต่อหน่วยไตมีปริมาณน้ำตาลมาก น้ำจึงถูกดูดกลับได้น้อย

กระดาษคำตอบ

ข้อ	ตัวเลือก				ระดับความมั่นใจ	
	ก	ข	ค	ง	มั่นใจ	ไม่มั่นใจ
5						

6. ผู้ป่วยรายหนึ่งมีความผิดปกติของระบบภูมิคุ้มกัน โดยแพทย์วินิจฉัยว่า ผู้ป่วยรายนี้มีความผิดปกติที่เซลล์เม็ดเลือดขาวชนิดฟาโกไซต์และเซลล์ที ต่อมาผู้ป่วยรายนี้ได้รับวัคซีน พบว่าไม่สามารถสร้างภูมิคุ้มกันต่อโรคนั้นได้

เพราะเหตุใดผู้ป่วยรายนี้จึงไม่สามารถสร้างภูมิคุ้มกันต่อโรคนั้นได้

- ก. เซลล์เม็ดเลือดขาวชนิดฟาโกไซต์ไม่สามารถจับกับแอนติบอดีได้
ข. เซลล์เม็ดเลือดขาวชนิดเซลล์ทีไม่สามารถส่งสัญญาณให้เซลล์บีแบ่งเซลล์ได้
ค. เซลล์เม็ดเลือดขาวชนิดเซลล์ทีไม่สามารถสร้างแอนติบอดีที่จำเพาะกับแอนติเจนได้
ง. เซลล์เม็ดเลือดขาวชนิดเซลล์ทีไม่สามารถกระตุ้นให้เซลล์บีจดจำและจำแนกแอนติเจนได้

กระดาษคำตอบ

ข้อ	ตัวเลือก				ระดับความมั่นใจ	
	ก	ข	ค	ง	มั่นใจ	ไม่มั่นใจ
6						

2. ข้อใดกล่าวถึงการรักษาคุณภาพของกรด-เบสในเลือดได้ถูกต้อง
- ถ้าเลือดเป็นเบส อัตราการหายใจจะลดลง เพื่อลดปริมาณ H^+
 - ถ้าเลือดเป็นเบส อัตราการหายใจจะเพิ่มขึ้น เพื่อเพิ่มปริมาณ H^+
 - ถ้าเลือดเป็นกรด อัตราการหายใจจะลดลง เพื่อเพิ่มปริมาณ CO_2
 - ถ้าเลือดเป็นกรด อัตราการหายใจจะเพิ่มขึ้น เพื่อลดปริมาณ CO_2

กระดาษคำตอบ

ข้อ	ตัวเลือก				ระดับความมั่นใจ			
	ก	ข	ค	ง	มั่นใจมาก	มั่นใจ	ไม่มั่นใจ	เดา
2								

องค์ประกอบที่ 2 การประเมินและออกแบบกระบวนการสืบเสาะหาความรู้ทางวิทยาศาสตร์

3. เมื่อบดเนื้อเยื่อของพืชหัวชนิดหนึ่งกับน้ำจะได้ของเหลวขาวขุ่นคล้ายน้ำแป้ง ถ้าต้องการทดสอบว่าของเหลวนั้นมีแป้งเป็นองค์ประกอบหรือไม่ ควรทำการทดลองโดยใช้หลอดไตบ่าง จากตารางการทดลองข้างล่างนี้

หลอดที่	ของเหลวจากพืชหัว	น้ำ	น้ำแป้งมัน	นมจืด	สารละลายไอโอดีน	สารละลายไบยูเรต
1	+	-	-	-	+	-
2	+	-	-	-	-	+
3	-	-	+	-	+	-
4	-	-	-	+	-	+
5	-	+	-	-	+	-
6	-	+	-	-	-	+

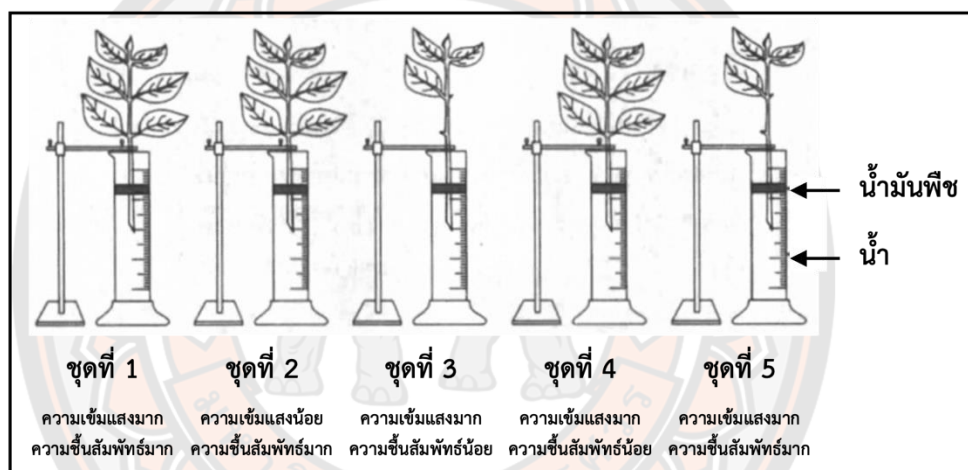
กำหนด: + มีการใช้สาร - ไม่มีการใช้สาร

- หลอดที่ 1 2
- หลอดที่ 2 6
- หลอดที่ 1 3 5
- หลอดที่ 2 4 6

กระดาษคำตอบ

ข้อ	ตัวเลือก				ระดับความมั่นใจ			
	ก	ข	ค	ง	มั่นใจมาก	มั่นใจ	ไม่มั่นใจ	เดา
3								

4. นักเรียนจัดชุดการทดลอง 5 ชุด โดยใช้กิ่งไม้ที่มีอายุเท่ากันจากต้นเดียวกัน เด็ดใบในชุดการทดลองที่ 3 และ 5 ออกบางส่วน จากนั้นแช่กิ่งไม้ในหลอดทดลองที่มีน้ำ 32 มิลลิลิตร และมีน้ำมันพืช 3 มิลลิลิตร เทห์บอยู่ แล้วตั้งทิ้งไว้ในสภาวะที่แตกต่างกัน ดังภาพ



เมื่อเวลาผ่านไป 30 นาที พบว่า ปริมาณน้ำคงเหลือในหลอดทดลอง ดังตาราง

ชุดการทดลองที่	ปริมาณน้ำคงเหลือในหลอดทดลอง (ml)
1	15
2	20
3	25
4	10
5	30

ข้อใดเลือกชุดการทดลองเพื่อศึกษาปัจจัยที่เกี่ยวข้องกับการคายน้ำและเปรียบเทียบอัตราการคายน้ำได้ถูกต้อง

	ชุดการทดลองที่ใช้	ปัจจัยที่ต้องการศึกษา	ผลการเปรียบเทียบอัตราการคายน้ำ
ก.	1 และ 2	ความเข้มแสง	ชุดการทดลองที่ 2 มีอัตราการคายน้ำมากกว่า ชุดการทดลองที่ 1
ข.	1 และ 4	ความชื้นสัมพัทธ์	ชุดการทดลองที่ 4 มีอัตราการคายน้ำมากกว่า ชุดการทดลองที่ 1
ค.	3 และ 4	จำนวนใบ	ชุดการทดลองที่ 3 มีอัตราการคายน้ำมากกว่า ชุดการทดลองที่ 4
ง.	3 และ 5	ความเข้มแสง	ชุดการทดลองที่ 5 มีอัตราการคายน้ำมากกว่า ชุดการทดลองที่ 3

กระดาษคำตอบ

ข้อ	ตัวเลือก				ระดับความมั่นใจ			
	ก	ข	ค	ง	มั่นใจมาก	มั่นใจ	ไม่มั่นใจ	เดา
4								

องค์ประกอบที่ 3 การแปลความหมายข้อมูลและการใช้ประจักษ์พยานในเชิงวิทยาศาสตร์

5. โรคเบาหวานชนิดที่ 1 เกิดจากตับอ่อนสร้าง “ฮอร์โมนอินซูลิน” ได้น้อย หรือสร้างไม่ได้เลย ซึ่งฮอร์โมนชนิดนี้ทำหน้าที่ช่วยให้ร่างกายสลายน้ำตาลมาใช้เป็นพลังงาน เมื่ออินซูลินในร่างกายไม่เพียงพอ ร่างกายไม่สามารถนำน้ำตาลไปใช้ได้ ทำให้ปริมาณน้ำตาลในเลือดสูง ร่างกายจะปรับตัวไปใช้พลังงานจากการสลายสารอาหารอื่น เช่น ไขมัน หรือ โปรตีน ซึ่งก่อให้เกิดการสะสมของของเสียที่เป็นอันตรายในเลือด น้ำตาลและของเสียปริมาณมากทำให้เลือดของผู้ป่วยโรคเบาหวานมีความเข้มข้นสูง ร่างกายจึงต้องขับน้ำตาลและของเสียผ่านทางปัสสาวะ ทำให้ผู้ป่วยโรคเบาหวานปัสสาวะบ่อยในปริมาณมากและรู้สึกกระหายน้ำ

ข้อใดไม่ใช่สาเหตุที่ทำให้ผู้ป่วยโรคเบาหวานปัสสาวะบ่อยในปริมาณมาก

- ก. น้ำจะออสโมซิสจากเซลล์มาสู่เลือด
- ข. ร่างกายกำจัดน้ำตาลที่มีมากเกินไปในเลือด

