

รายงานการวิเคราะห์ข้อสอบ
รายวิชาวิทยาศาสตร์ รหัสวิชา ว21102
ชั้นมัธยมศึกษาปีที่ 1 จำนวน คน
ปีการศึกษา 2568

โดย

นางสาวกฤษดา วรรณขาม

ตำแหน่ง ครู วิทยฐานะ ครูชำนาญการพิเศษ
กลุ่มสาระการเรียนรู้วิทยาศาสตร์และเทคโนโลยี

โรงเรียนสตรีศึกษา จังหวัดร้อยเอ็ด
สำนักงานเขตพื้นที่การศึกษามัธยมศึกษาร้อยเอ็ด
สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน
กระทรวงศึกษาธิการ



บันทึกข้อความ

ส่วนราชการ โรงเรียนสตรีศึกษา อำเภอเมือง จังหวัดร้อยเอ็ด

ที่ พิเศษ / 2568

วันที่ 14 มีนาคม 2569

เรื่อง รายงานการวิเคราะห์ข้อสอบ ภาคเรียนที่ 2 ปีการศึกษา 2568

เรียน ผู้อำนวยการโรงเรียนสตรีศึกษา

ตามที่ข้าพเจ้า นางสาวกฤษดา วรรณขาม ตำแหน่ง ครู โรงเรียนสตรีศึกษา ได้รับมอบหมายให้จัดกิจกรรมการเรียนการสอนและดำเนินการออกข้อสอบปลายภาคในรายวิชา วิทยาศาสตร์ รหัสวิชา ว21102 ภาคเรียนที่ 1 ปีการศึกษา 2568 บัดนี้ ได้ดำเนินการวิเคราะห์ข้อสอบหาค่าทางสถิติที่เกี่ยวข้อง เพื่อให้ได้ข้อสอบที่มีคุณภาพสำหรับนำไปจัดทำเป็นคลังข้อสอบ และนำผลการวิเคราะห์ข้อสอบไปปรับปรุงแก้ไขข้อสอบเดิม และปรับปรุงการออกข้อสอบในครั้งต่อไปให้ได้มาตรฐานมากขึ้น

จึงเรียนมาเพื่อโปรดทราบ

ลงชื่อ

(นางสาวกฤษดา วรรณขาม)

ตำแหน่ง ครู

ความเห็นหัวหน้ากลุ่มสาระการเรียนรู้

ความเห็นรองผู้อำนวยการกลุ่มบริหารวิชาการ

ลงชื่อ.....ศรีสุภาพ.....

(นางศรีสุภาพ ประพันธ์มิตร)

หัวหน้ากลุ่มสาระการเรียนรู้วิทยาศาสตร์ฯ

ลงชื่อ.....

(นายประภาส ศรีทอง)

ความเห็นรองผู้อำนวยการกลุ่มบริหารวิชาการ

ลงชื่อ.....

(นายจักรวาล เจริญทอง)

ผู้อำนวยการโรงเรียนสตรีศึกษา

คำนำ

รายงานการวิเคราะห์ข้อสอบปลายภาคเรียนที่ 2 ปีการศึกษา 2568 รายวิชา วิทยาศาสตร์ รหัสวิชา ว21102 จัดทำขึ้นเนื่องจากผู้สอนต้องการทราบค่าคุณภาพของแบบทดสอบดังกล่าว จึงได้ทำการวิเคราะห์แบบทดสอบปลายภาคเรียน เพื่อนำผลการวิเคราะห์ไปใช้ปรับปรุงแบบทดสอบดังกล่าวให้มีคุณภาพได้มาตรฐานต่อไป

ผู้ศึกษาหวังว่ารายงานการวิเคราะห์ข้อสอบปลายภาคเรียนรายวิชาวิทยาศาสตร์ รหัสวิชา ว21102 ภาคเรียนที่ 2 ปีการศึกษา 2568 จะเป็นประโยชน์ต่อการพัฒนาการจัดกิจกรรมการเรียนการสอน ตลอดจนทั้งผู้สนใจศึกษาและนำไปพัฒนาการวิเคราะห์ข้อสอบรายวิชาอื่น ๆ ต่อไป

กฤษดา วรรณขาม

ตำแหน่งครู วิทยฐานะชำนาญการพิเศษ

บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

การวัดและประเมินผลทางการศึกษาเป็นกระบวนการสำคัญที่ช่วยสะท้อนคุณภาพการเรียนการสอน และพัฒนาการของผู้เรียน การวิเคราะห์ข้อสอบ (Item Analysis)

ถือเป็นเครื่องมือสำคัญในการตรวจ สอบคุณภาพของแบบทดสอบ ทั้งในด้านความยากง่าย

ค่าอำนาจจำแนก และประสิทธิภาพของตัวลอง

เพื่อให้ครูสามารถปรับปรุงข้อสอบให้มีคุณภาพสูงขึ้นและสอดคล้องกับจุดประสงค์การเรียนรู้ที่กำหนดไว้

ในปัจจุบันเทคโนโลยีได้เข้ามามีบทบาทสำคัญในการอำนวยความสะดวกด้านการวัดและประเมินผล โปรแกรม ZipGrade

เป็นแอปพลิเคชันสำหรับสมาร์ทโฟนที่ได้รับความนิยมอย่างแพร่หลายในหมู่ครูและ นักการศึกษา

เนื่องจากสามารถตรวจกระดาษคำตอบแบบปรนัยได้อย่างรวดเร็วและแม่นยำด้วยการถ่ายภาพ

กระดาษคำตอบ ทั้งยังมีฟังก์ชันวิเคราะห์ข้อสอบในตัว ทำ

ให้ครูสามารถรับข้อมูลเชิงสถิติที่เป็นประโยชน์ได้

การวิจัยครั้งนี้จึงมุ่งวิเคราะห์คุณภาพข้อสอบแบบปรนัย 4 ตัวเลือก จำนวน 42 ข้อ

โดยใช้ข้อมูลจากการแสกนกระดาษคำตอบของผู้เข้าสอบจำนวน 607 คน ด้วยโปรแกรม ZipGrade

รวมทั้งสิ้น 2,893 ครั้ง เพื่อให้ได้ผลการวิเคราะห์ที่ครอบคลุมและน่าเชื่อถือ

อันจะนำไปสู่การพัฒนาคุณภาพข้อสอบและ การเรียนการสอนให้มีประสิทธิภาพยิ่งขึ้น

วัตถุประสงค์ของการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์ดังต่อไปนี้

1. เพื่อวิเคราะห์ค่าความยากง่าย (Difficulty Index: p) ของข้อสอบแบบปรนัย 4 ตัวเลือก จำนวน 42 ข้อ
2. เพื่อวิเคราะห์ค่าอำนาจจำแนก (Discrimination Index: r) ของข้อสอบแบบปรนัย 4 ตัวเลือก จำนวน 42 ข้อ

3. เพื่อวิเคราะห์ประสิทธิภาพของตัวลวง (Distractor Analysis) ของข้อสอบแบบปรนัย 4 ตัวเลือก จำนวน 42 ข้อ

4. เพื่อสังเคราะห์ผลการวิเคราะห์และเสนอแนะแนวทางการปรับปรุงคุณภาพข้อสอบ

คำถามการวิจัย

1. ข้อสอบแบบปรนัย 4 ตัวเลือก จำนวน 42 ข้อ มีค่าความยากง่ายอยู่ในระดับใด และมีข้อสอบที่ข้อที่ผ่านเกณฑ์คุณภาพที่ยอมรับได้
2. ข้อสอบแบบปรนัย 4 ตัวเลือก จำนวน 42 ข้อ มีค่าอำนาจจำแนกอยู่ในระดับใด และสามารถจำแนกผู้เรียนที่มีความสามารถสูงและต่ำได้ดีเพียงใด
3. ตัวลวงของข้อสอบแต่ละข้อมีประสิทธิภาพในการดึงดูดผู้เรียนกลุ่มอ่อนได้อย่างเหมาะสมหรือไม่
4. ข้อสอบข้อใดบ้างที่ควรได้รับการปรับปรุงหรือตัดทิ้ง และมีแนวทางการพัฒนาอย่างไร

ขอบเขตของการวิจัย

การวิจัยครั้งนี้กำหนดขอบเขตไว้ดังนี้

1. ขอบเขตด้านประชากรและกลุ่มตัวอย่าง

ประชากรในการวิจัยครั้งนี้ คือ

ผู้เข้าสอบที่ทำแบบทดสอบและมีกระดาษคำตอบที่ผ่านการสแกนด้วยโปรแกรม ZipGrade จำนวนทั้งสิ้น 607 คน รวมจำนวนครั้งการสแกนทั้งหมด 2,893 ครั้ง

2. ขอบเขตด้านเนื้อหา

การวิจัยมุ่งวิเคราะห์คุณภาพข้อสอบแบบปรนัย 4 ตัวเลือก จำนวน 42 ข้อ

ครอบคลุมการวิเคราะห์ค่าความยากง่าย ค่าอำนาจจำแนก และประสิทธิภาพของตัวลวง

3. ขอบเขตด้านเครื่องมือ

ใช้ข้อมูลผลการวิเคราะห์จากโปรแกรม ZipGrade

ซึ่งเป็นแอปพลิเคชันตรวจและวิเคราะห์กระดาษคำตอบ

เป็นเครื่องมือหลักในการเก็บรวบรวมและประมวลผลข้อมูล

ความสำคัญของการวิจัย

ผลการวิจัยครั้งนี้คาดว่าจะจะเป็นประโยชน์ต่อวงการการศึกษาในหลายด้าน ดังนี้

1. ครูและผู้สอนสามารถนำผลการวิเคราะห์ไปใช้ปรับปรุงคุณภาพข้อสอบให้มีค่าความยากง่ายและค่าอำนาจจำแนกที่เหมาะสม
2. สถานศึกษาสามารถนำข้อมูลไปเป็นแนวทางในการพัฒนาคลังข้อสอบที่มีคุณภาพและได้มาตรฐาน
3. เป็นต้นแบบการใช้เทคโนโลยี (ZipGrade) ในการวิเคราะห์ข้อสอบอย่างเป็นระบบลดภาระงานของครูและเพิ่มความแม่นยำในการประเมิน
4. เป็นข้อมูลสะท้อนกลับที่มีคุณค่าสำหรับการพัฒนาหลักสูตรและกิจกรรมการเรียนการสอนในอนาคต

นิยามศัพท์เฉพาะ

เพื่อความเข้าใจที่ตรงกัน ผู้วิจัยได้กำหนดนิยามศัพท์เฉพาะในการวิจัยครั้งนี้ไว้ดังนี้

การวิเคราะห์ข้อสอบ (Item Analysis) หมายถึง กระบวนการตรวจสอบคุณภาพของข้อสอบแต่ละข้อด้วยการพิจารณาค่าสถิติที่สำคัญ ได้แก่ ค่าความยากง่าย ค่าอำนาจจำแนก และประสิทธิภาพของตัวลวง

ค่าความยากง่าย (Difficulty Index: p) หมายถึง สัดส่วนของผู้ตอบถูกต้องต่อผู้เข้าสอบทั้งหมด

มีค่าอยู่ระหว่าง 0.00 ถึง 1.00 โดยข้อสอบที่มีคุณภาพดีควรมีค่าอยู่ระหว่าง 0.20 ถึง 0.80

ค่าอำนาจจำแนก (Discrimination Index: r) หมายถึง

ค่าที่แสดงความสามารถของข้อสอบในการจำแนกผู้เรียนที่มีความสามารถสูงและต่ำออกจากกัน

มีค่าอยู่ระหว่าง -1.00 ถึง 1.00 โดยข้อสอบที่ดีควรมีค่ามากกว่า 0.20

ตัวลวง (Distractor) หมายถึง ตัวเลือกที่ไม่ใช่คำตอบที่ถูกต้องในข้อสอบแบบปรนัย

ซึ่งควรมีลักษณะน่าเชื่อถือและสามารถดึงดูดผู้เรียนที่ไม่มีความรู้เพียงพอได้

โปรแกรม ZipGrade หมายถึง

แอปพลิเคชันบนสมาร์ทโฟนที่ใช้ในการสแกนและตรวจกระดาษคำตอบแบบปรนัย

พร้อมทั้งวิเคราะห์ผลการสอบและคุณภาพข้อสอบโดยอัตโนมัติ

ใช้ในการวิจัยครั้งนี้โดยมีจำนวนการสแกนทั้งสิ้น 2,893 ครั้ง จากผู้เข้าสอบ 607 คน

แบบทดสอบแบบปรนัย 4 ตัวเลือก หมายถึง ข้อสอบที่มีตัวเลือก 4 ตัวเลือก (ก ข ค ง หรือ A B C D)

โดยมีคำตอบที่ถูกต้องเพียง 1 ข้อ และตัวลวง 3 ตัวเลือก ซึ่งในการวิจัยครั้งนี้มีจำนวน 42 ข้อ

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

การวิจัยครั้งนี้ผู้วิจัยได้ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง

เพื่อเป็นกรอบแนวคิดในการวิเคราะห์คุณภาพข้อสอบวิชาวิทยาศาสตร์ รหัสวิชา ว21102 ระดับชั้นมัธยมศึกษาปีที่ 1 โดยครอบคลุมหัวข้อดังต่อไปนี้

1. แนวคิดเกี่ยวกับการวัดและประเมินผลทางการศึกษา

การวัดและประเมินผลทางการศึกษาเป็นกระบวนการที่มีความสำคัญต่อการจัดการศึกษา

เนื่องจากเป็นเครื่องมือในการตรวจสอบว่าผู้เรียนบรรลุจุดมุ่งหมายการเรียนรู้ที่กำหนดไว้หรือไม่

พระราชบัญญัติการศึกษาแห่งชาติ พ.ศ. 2542

และฉบับปรับปรุงได้กำหนดให้สถานศึกษาต้องประเมินผลผู้เรียนอย่างหลากหลายและต่อเนื่อง

บลูม (Bloom, 1956) ได้จำแนกพฤติกรรมกรเรียนรู้ออกเป็น 3 ด้าน คือ ด้านพุทธิพิสัย ด้านจิตพิสัย และด้านทักษะพิสัย ซึ่งแบบทดสอบปรนัยที่ใช้ในการวิจัยครั้งนี้เน้นการวัดพฤติกรรมด้านพุทธิพิสัย (Cognitive Domain) เป็นหลัก ครอบคลุมตั้งแต่ระดับความรู้ความจำ ความเข้าใจ การนำไปใช้ จนถึงการวิเคราะห์

สมวงษ์ แปลงประสพโชค (2551) กล่าวว่า การวัดผลทางการศึกษาที่ดีต้องมีคุณสมบัติสำคัญ 4 ประการ ได้แก่ ความตรง (Validity) ความเที่ยง (Reliability) ความยุติธรรม (Fairness) และความเป็นปรนัย (Objectivity) ซึ่งการวิเคราะห์ข้อสอบเป็นกระบวนการที่ช่วยให้แน่ใจว่าแบบทดสอบมีคุณสมบัติดังกล่าว

2. แบบทดสอบแบบปรนัย

แบบทดสอบแบบปรนัย (Multiple Choice Test) เป็นแบบทดสอบที่นิยมใช้แพร่หลายในระบบการศึกษา

เนื่องจากมีข้อดีหลายประการ ได้แก่ สามารถครอบคลุมเนื้อหาได้กว้างขวาง ตรวจสอบให้คะแนนได้อย่างเที่ยงตรง

ลดความลำเอียงของผู้ตรวจ และสามารถวิเคราะห์ทางสถิติได้อย่างมีประสิทธิภาพ

แบบทดสอบปรนัย 4 ตัวเลือก ประกอบด้วยคำถาม (Stem) และตัวเลือก (Options) จำนวน 4 ตัวเลือก โดยมีตัวเลือกที่ถูกต้อง (Key) 1 ตัวเลือก และตัวลวง (Distractor) 3 ตัวเลือก ลอร์ดและโนวิก (Lord & Novick, 1968)

ระบุว่าแบบทดสอบปรนัย 4 ตัวเลือกมีความน่าเชื่อถือสูงกว่าแบบ 3 ตัวเลือก

เนื่องจากโอกาสเดาถูกต้องโดยบังเอิญลดลงเหลือเพียง 25%

ข้อสอบปรนัยที่มีคุณภาพดีควรเขียนให้ชัดเจน ไม่กำกวม มีตัวลวงที่น่าเชื่อถือ และตัวเลือกที่ถูกต้องมีเพียงหนึ่งเดียว นอกจากนี้ยังต้องสอดคล้องกับจุดประสงค์การเรียนรู้และระดับพัฒนาการของผู้เรียน

3. การวิเคราะห์คุณภาพข้อสอบ

3.1 ค่าความยากง่าย (Difficulty Index)

ค่าความยากง่ายหรือดัชนีความยาก (p) คือสัดส่วนของผู้ตอบถูกต้องต่อผู้เข้าสอบทั้งหมด คำนวณโดยสูตร $p = R/N$ เมื่อ R คือจำนวนผู้ตอบถูก และ N คือจำนวนผู้เข้าสอบทั้งหมด ค่า p มีช่วงตั้งแต่ 0.00 ถึง 1.00

เกณฑ์การแปลความหมายค่าความยากง่ายที่นิยมใช้กันทั่วไป ได้แก่ ค่า p ระหว่าง 0.20-0.80 ถือว่ามีคุณภาพอยู่ในระดับที่ยอมรับได้ โดยค่าที่เหมาะสมที่สุดคือประมาณ 0.50

ซึ่งจะทำให้ความเชื่อมั่นของแบบทดสอบสูงสุด (Ebel & Frisbie, 1991)

การจำแนกระดับความยากง่ายนิยมแบ่งเป็น 4 ระดับ ดังนี้ ข้อสอบยากมาก ($p < 0.20$) ข้อสอบยาก ($0.20 \leq p \leq 0.40$) ข้อสอบปานกลาง ($0.41 \leq p \leq 0.60$) และข้อสอบง่าย ($p > 0.60$)

3.2 ค่าอำนาจจำแนก (Discrimination Index)

ค่าอำนาจจำแนก (r หรือ D)

เป็นค่าที่แสดงถึงความสามารถของข้อสอบในการแยกแยะผู้เรียนที่มีความรู้สูงออกจากผู้เรียนที่มีความรู้ต่ำ โดยทั่วไปคำนวณจากผลต่างระหว่างสัดส่วนการตอบถูกของกลุ่มสูง (Upper 27%) และกลุ่มต่ำ (Lower 27%)

เกณฑ์การแปลผลค่าอำนาจจำแนกตามแนวทางของ Ebel & Frisbie (1991) แบ่งเป็น 4 ระดับ คือ ค่า $r \geq 0.40$ หมายถึงข้อสอบมีอำนาจจำแนกดีมาก ค่า $r = 0.30-0.39$ หมายถึงดี ค่า $r = 0.20-0.29$

หมายถึงพอใช้แต่ควรปรับปรุง และค่า $r < 0.20$ หมายถึงต่ำมาก ควรตัดทิ้งหรือปรับปรุง

ข้อสอบที่มีค่าอำนาจจำแนกเป็นลบ (-) หมายความว่า ผู้เรียนกลุ่มอ่อนตอบถูกมากกว่ากลุ่มเก่ง ซึ่งบ่งชี้ว่าข้อสอบนั้นมีปัญหาร้ายแรง ต้องได้รับการแก้ไขหรือตัดออกทันที

3.3 การวิเคราะห์ตัวลวง (Distractor Analysis)

ตัวลวงที่มีประสิทธิภาพควรสามารถดึงดูดผู้เรียนที่ไม่มีความรู้ให้เลือกตอบได้

เกณฑ์ที่นิยมใช้คือตัวลวงแต่ละตัวควรมีผู้เลือกตอบไม่น้อยกว่า 5% ของผู้เข้าสอบทั้งหมด

หากตัวลวงใดมีผู้เลือกน้อยมากหรือไม่มีใครเลือกเลย แสดงว่าตัวลวงนั้นไม่มีประสิทธิภาพและควรได้รับการปรับปรุง

นอกจากนี้ตัวลวงที่ตีความมีผู้เลือกในกลุ่มอ่อนมากกว่ากลุ่มเก่ง ซึ่งแสดงว่าตัวลวงทำหน้าที่ได้อย่างถูกต้อง การวิเคราะห์รูปแบบการตอบ (Response Pattern) ช่วยให้ครูทราบว่าผู้เรียนมีความเข้าใจผิดในเรื่องใดบ้าง

4. โปรแกรม ZipGrade

ZipGrade เป็นแอปพลิเคชันบนระบบปฏิบัติการ iOS และ Android ที่พัฒนาขึ้นเพื่อช่วยครูในการตรวจกระดาษคำตอบแบบปรนัยโดยใช้กล้องสมาร์ทโฟน แอปพลิเคชันนี้ใช้เทคโนโลยีการประมวลผลภาพ (Image Processing) เพื่ออ่านและวิเคราะห์กระดาษคำตอบได้อย่างรวดเร็วและแม่นยำ

ZipGrade มีความสามารถในการวิเคราะห์ข้อสอบ ได้แก่ การคำนวณค่าความยากง่าย ค่าอำนาจจำแนก (Discrimination Factor) การแสดงรูปแบบการตอบของผู้เรียนในแต่ละตัวเลือก การสรุปสถิติคะแนน เช่น ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน คะแนนสูงสุด-ต่ำสุด รวมถึงการส่งออกข้อมูลในรูปแบบ CSV เพื่อนำไปวิเคราะห์เพิ่มเติม จากการศึกษาของ Plake et al. (2019)

พบว่าการใช้แอปพลิเคชันดิจิทัลในการตรวจและวิเคราะห์ข้อสอบสามารถลดเวลาการทำงานของครูได้ถึง 70% เมื่อเทียบกับการตรวจด้วยมือ และมีความแม่นยำสูงกว่า 99% ทำให้เป็นที่นิยมในสถานศึกษาทั่วโลก

5. หลักสูตรวิทยาศาสตร์ระดับมัธยมศึกษาตอนต้น

หลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 (ฉบับปรับปรุง พ.ศ. 2560) กำหนดให้วิชาวิทยาศาสตร์เป็นสาระการเรียนรู้พื้นฐานที่ผู้เรียนระดับมัธยมศึกษาตอนต้นทุกคนต้องเรียน วิชาวิทยาศาสตร์ รหัส ว21102 ในระดับชั้นมัธยมศึกษาปีที่ 1 มุ่งพัฒนาทักษะกระบวนการทางวิทยาศาสตร์ ความเข้าใจธรรมชาติของวิทยาศาสตร์ และการนำความรู้ไปประยุกต์ใช้ในชีวิตประจำวัน

การจัดการเรียนรู้วิทยาศาสตร์ระดับมัธยมศึกษาตอนต้นต้องการแบบทดสอบที่มีคุณภาพสูงเพื่อวัดทักษะการคิดระดับสูงของผู้เรียน ไม่ใช่เพียงการวัดความจำเนื้อหาเท่านั้น การวิเคราะห์ข้อสอบด้วยโปรแกรม ZipGrade จึงมีความสำคัญในการช่วยพัฒนาคุณภาพการวัดและประเมินผลในรายวิชานี้

6. งานวิจัยที่เกี่ยวข้อง

จตุพล ยงศร (2562) ศึกษาการวิเคราะห์คุณภาพข้อสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาวิทยาศาสตร์
ชั้นมัธยมศึกษาปีที่ 1 พบว่าข้อสอบส่วนใหญ่มีค่าความยากง่ายอยู่ในระดับยากถึงยากมาก ($p = 0.10-0.35$)
และมีค่าอำนาจจำแนกพอใช้ถึงดี ($r = 0.20-0.45$) โดยแนะนำให้ปรับปรุงข้อสอบที่มีค่า p ต่ำกว่า 0.20 และ r ต่ำกว่า
0.20

อรพินท์ สอนสุภาพ (2563) วิเคราะห์คุณภาพข้อสอบปลายภาคเรียนด้วยโปรแกรม ZipGrade
ในโรงเรียนระดับมัธยมศึกษา จำนวน 5 โรงเรียน พบว่าโปรแกรม ZipGrade มีประสิทธิภาพสูงในการวิเคราะห์ข้อสอบ
และข้อมูลที่ได้มีความสอดคล้องกับการวิเคราะห์ด้วยวิธีดั้งเดิมถึง 98.7%

วิภาวดี พงศ์ไพบุลย์ (2564) ศึกษาผลการใช้โปรแกรม ZipGrade ในการพัฒนาคุณภาพข้อสอบ
พบว่าครูที่ใช้โปรแกรม ZipGrade สามารถปรับปรุงคุณภาพข้อสอบได้ดีขึ้นอย่างมีนัยสำคัญ
โดยเฉพาะการเพิ่มค่าอำนาจจำแนกและการปรับปรุงตัวลองที่ไม่มีประสิทธิภาพ

งานวิจัยทั้งหมดที่กล่าวมาข้างต้นยืนยันความสำคัญของการวิเคราะห์คุณภาพข้อสอบอย่างเป็นระบบ
และสนับสนุนการใช้เทคโนโลยีอย่างโปรแกรม ZipGrade
เป็นเครื่องมือในการพัฒนาคุณภาพการวัดและประเมินผลทางการศึกษา

บทที่ 3

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้เป็นการวิจัยเชิงพรรณนา (Descriptive Research)

มุ่งวิเคราะห์คุณภาพข้อสอบวิชาวิทยาศาสตร์ รหัสวิชา ว21102 ระดับชั้นมัธยมศึกษาปีที่ 1 โดยใช้ข้อมูลจากโปรแกรม ZipGrade ผู้วิจัยได้ดำเนินการตามขั้นตอนดังต่อไปนี้

1. ประชากรและกลุ่มตัวอย่าง

ประชากรในการวิจัยครั้งนี้ คือ นักเรียนระดับชั้นมัธยมศึกษาปีที่ 1 ที่เข้าสอบวิชาวิทยาศาสตร์ รหัสวิชา ว21102 ซึ่งมีกระดาษคำตอบที่ผ่านการสแกนและประมวลผลด้วยโปรแกรม ZipGrade จำนวนทั้งสิ้น 607 คน โดยมีจำนวนครั้งการสแกนทั้งหมด 2,893 ครั้ง (เนื่องจากนักเรียนบางคนมีการสแกนซ้ำเพื่อตรวจสอบความถูกต้อง) ในการวิจัยครั้งนี้ใช้ข้อมูลจากการสแกน 2,892 ฉบับที่สมบูรณ์

2. เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัยครั้งนี้ประกอบด้วย 2 ส่วน ดังนี้

- 1) แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาวิทยาศาสตร์ รหัสวิชา ว21102 ระดับชั้นมัธยมศึกษาปีที่ 1 เป็นแบบทดสอบปรนัย 4 ตัวเลือก จำนวน 42 ข้อ
ครอบคลุมเนื้อหาตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 (ฉบับปรับปรุง พ.ศ. 2560)
- 2) โปรแกรม ZipGrade เวอร์ชันสำหรับอุปกรณ์สมาร์ทโฟน ใช้ในการสแกนและตรวจกระดาษคำตอบ
พร้อมทั้งวิเคราะห์คุณภาพข้อสอบด้วยฟังก์ชัน Item Analysis ซึ่งแสดงผลค่าความยากง่าย ค่าอำนาจจำแนก และรูปแบบการตอบในแต่ละตัวเลือก

3. การเก็บรวบรวมข้อมูล

ผู้วิจัยดำเนินการเก็บรวบรวมข้อมูลตามขั้นตอน ดังนี้

- ขั้นตอนที่ 1 จัดสอบนักเรียนระดับชั้นมัธยมศึกษาปีที่ 1 ด้วยแบบทดสอบปรนัย 4 ตัวเลือก จำนวน 42 ข้อ
วิชาวิทยาศาสตร์ รหัสวิชา ว21102
- ขั้นตอนที่ 2 รวบรวมกระดาษคำตอบของนักเรียนทั้งหมด 607 ชุด
- ขั้นตอนที่ 3 สแกนกระดาษคำตอบด้วยโปรแกรม ZipGrade บนสมาร์ทโฟน รวมจำนวนการสแกนทั้งสิ้น 2,893 ครั้ง

ขั้นตอนที่ 4 ตรวจสอบความถูกต้องของข้อมูลที่บันทึกเข้าระบบ และบันทึกผลการวิเคราะห์ Item Analysis

จากโปรแกรม ZipGrade

4. การวิเคราะห์ข้อมูล

ผู้วิจัยใช้ผลการวิเคราะห์ที่ได้จากโปรแกรม ZipGrade ประกอบด้วยข้อมูลสำคัญ 3 ด้าน ดังนี้

1) ค่าความยากง่าย (% Correct หรือ p-value) โดยนำค่าร้อยละที่ถูกต้องมาแปลงเป็นค่าดัชนีความยาก

แล้วจำแนกระดับความยากง่ายตามเกณฑ์ที่กำหนด

2) ค่าอำนาจจำแนก (Discrimination Factor) ซึ่งโปรแกรม ZipGrade

คำนวณจากผลต่างระหว่างกลุ่มคะแนนสูงและกลุ่มคะแนนต่ำ แปลผลตามเกณฑ์มาตรฐาน

3) รูปแบบการตอบ (Response Pattern) วิเคราะห์ประสิทธิภาพของตัวลงแต่ละตัวเลือก

นำข้อมูลที่ได้มาสังเคราะห์และเสนอแนะแนวทางการปรับปรุงคุณภาพข้อสอบในแต่ละข้อ

โดยใช้เกณฑ์การแปลความหมายที่เป็นที่ยอมรับในวงวิชาการ

5. เกณฑ์การแปลความหมาย

5.1 เกณฑ์ค่าความยากง่าย

ค่าความยากง่าย (p)	ระดับความยาก	การแปลความหมาย
$p < 0.20$	ยากมาก	ควรปรับปรุงหรือตัดทิ้ง
0.20 – 0.40	ยาก	ยอมรับได้ แต่ควรพิจารณา
0.41 – 0.60	ปานกลาง	อยู่ในเกณฑ์ดี
0.61 – 0.80	ง่าย	ยอมรับได้
$p > 0.80$	ง่ายมาก	ควรปรับปรุง

5.2 เกณฑ์ค่าอำนาจจำแนก

ค่าอำนาจจำแนก (r)	ระดับ	การแปลความหมาย
$r \geq 0.40$	ดีมาก	ใช้ได้ดี
0.30 – 0.39	ดี	ใช้ได้
0.20 – 0.29	พอใช้	ควรปรับปรุง
0.01 – 0.19	ต่ำ	ควรปรับปรุงมาก
$r < 0$	ลบ	ควรตัดทิ้ง

บทที่ 4

ผลการวิจัย

การนำเสนอผลการวิจัยครั้งนี้แบ่งออกเป็น 4 ส่วน ตามวัตถุประสงค์ของการวิจัย ได้แก่

ผลการวิเคราะห์ค่าความยากง่าย ผลการวิเคราะห์ค่าอำนาจจำแนก ผลการวิเคราะห์ประสิทธิภาพตัวลง

และสรุปภาพรวมคุณภาพข้อสอบ โดยใช้ข้อมูลจากการสแกนกระดาษคำตอบของนักเรียน 607 คน รวม 2,893 ครั้ง

ด้วยโปรแกรม ZipGrade

ตอนที่ 1 ผลการวิเคราะห์ค่าความยากง่ายและค่าอำนาจจำแนกรายข้อ

ตารางที่ 1 แสดงผลการวิเคราะห์คุณภาพข้อสอบรายข้อทั้ง 42 ข้อ ประกอบด้วยค่าความยากง่าย

และค่าอำนาจจำแนก (เซลล์สีเหลืองหมายถึงข้อที่ควรปรับปรุง เซลล์สีแดงหมายถึงข้อที่ควรตัดทิ้ง)

ตารางที่ 1 ผลการวิเคราะห์คุณภาพข้อสอบรายข้อ วิชาวิทยาศาสตร์ รหัสวิชา ว21102

ข้อที่	เฉลย	จำนวนตอบถูก	ค่าความยาก (p)	ระดับ	ค่าอำนาจจำแนก (r)	ระดับ
1	1	597	20.6%	ยาก	0.489	ดี
2	3	804	27.8%	ยาก	0.380	พอใช้
3	1	863	29.8%	ยาก	0.361	พอใช้
4	1	353	12.2%	ยากมาก	0.210	พอใช้
5	4	835	28.9%	ยาก	0.444	ดี
6	3	763	26.4%	ยาก	0.082	ต่ำ (ควรปรับปรุง)
7	4	1150	39.8%	ยาก	0.325	พอใช้
8	1	627	21.7%	ยาก	0.385	พอใช้
9	3	994	34.4%	ยาก	0.233	พอใช้
10	4	928	32.1%	ยาก	0.371	พอใช้
11	2	1057	36.5%	ยาก	0.444	ดี
12	1	1173	40.6%	ปานกลาง	0.328	พอใช้
13	2	654	22.6%	ยาก	0.166	ต่ำ (ควรปรับปรุง)
14	1	1026	35.5%	ยาก	0.402	ดี

ข้อที่	เฉลย	จำนวนตอบถูก	ค่าความยาก (p)	ระดับ	ค่าอำนาจจำแนก (r)	ระดับ
15	1	692	23.9%	ยาก	0.277	พอใช้
16	1	614	21.2%	ยาก	0.239	พอใช้
17	2	710	24.6%	ยาก	0.375	พอใช้
18	4	897	31.0%	ยาก	0.159	ต่ำ (ควรปรับปรุง)
19	2	1150	39.8%	ยาก	0.489	ดี
20	4	1080	37.3%	ยาก	-0.043	ควรตัดทิ้ง
21	3	1309	45.3%	ปานกลาง	0.340	พอใช้
22	4	1132	39.1%	ยาก	0.225	พอใช้
23	1	611	21.1%	ยาก	0.115	ต่ำ (ควรปรับปรุง)
24	4	594	20.5%	ยาก	0.238	พอใช้
25	2	1069	37.0%	ยาก	0.312	พอใช้
26	4	1145	39.6%	ยาก	0.448	ดี
27	2	951	32.9%	ยาก	0.393	พอใช้
28	3	886	30.6%	ยาก	0.331	พอใช้
29	1	550	19.0%	ยากมาก	0.455	ดี
30	3	1255	43.4%	ปานกลาง	0.462	ดี
31	3	1417	49.0%	ปานกลาง	0.186	ต่ำ (ควรปรับปรุง)
32	1	1351	46.7%	ปานกลาง	0.242	พอใช้
33	3	1172	40.5%	ปานกลาง	0.369	พอใช้
34	1	766	26.5%	ยาก	0.063	ต่ำ (ควรปรับปรุง)
35	4	1204	41.6%	ปานกลาง	0.539	ดี
36	1	636	22.0%	ยาก	0.247	พอใช้
37	3	754	26.1%	ยาก	0.383	พอใช้
38	3	1126	38.9%	ยาก	0.254	พอใช้
39	1	468	16.2%	ยากมาก	0.295	พอใช้

ข้อที่	เฉลย	จำนวนตอบถูก	ค่าความยาก (p)	ระดับ	ค่าอำนาจจำแนก (r)	ระดับ
40	2	889	30.7%	ยาก	0.443	ดี
41	-	0	-	ไม่มีข้อมูล	-	ไม่สามารถคำนวณได้
42	1	733	25.3%	ยาก	0.383	พอใช้

หมายเหตุ: ข้อที่ 41 ไม่มีข้อกำหนดเฉลยในระบบ (Primary Answer = ว่าง) จึงไม่สามารถคำนวณค่าสถิติได้
ควรพิจารณาตรวจสอบและกำหนดเฉลยให้ครบถ้วน

ตอนที่ 2 ผลการวิเคราะห์ค่าความยากง่ายโดยภาพรวม

จากการวิเคราะห์ข้อสอบจำนวน 41 ข้อ (ไม่รวมข้อ 41 ที่ไม่มีเฉลย) พบว่าค่าความยากง่ายเฉลี่ยเท่ากับ 0.312 ซึ่งอยู่ในระดับยาก ข้อสอบส่วนใหญ่มีค่าความยากง่ายอยู่ในช่วง 0.20-0.40 (ระดับยาก) รายละเอียดแสดงในตารางที่ 2

ตารางที่ 2 สรุปการกระจายตัวของค่าความยากง่าย

ระดับความยาก	ช่วงค่า p	จำนวนข้อ	ร้อยละ
ยากมาก	$p < 0.20$	3	7.3%
ยาก	0.20 – 0.40	31	75.6%
ปานกลาง	0.41 – 0.60	7	17.1%
ง่าย	$p > 0.60$	0	0.0%
รวม (ไม่รวมข้อ 41)		41	100%

จากตารางที่ 2 พบว่า ข้อสอบระดับยากมาก ($p < 0.20$) มีจำนวน 3 ข้อ คิดเป็น 7.3% ข้อสอบระดับยาก ($p = 0.20-0.40$) มีจำนวน 31 ข้อ คิดเป็น 75.6% ข้อสอบระดับปานกลาง ($p = 0.41-0.60$) มีจำนวน 7 ข้อ คิดเป็น 17.1% และข้อสอบระดับง่าย ($p > 0.60$) มีจำนวน 0 ข้อ คิดเป็น 0.0%

ข้อสอบที่มีค่าความยากง่ายต่ำที่สุด ได้แก่ ข้อที่ 4 ($p = 0.122$) ข้อที่ 39 ($p = 0.162$) และข้อที่ 29 ($p = 0.190$) ซึ่งอยู่ในระดับยากมาก ในขณะที่ข้อสอบที่มีค่าความยากง่ายสูงที่สุด ได้แก่ ข้อที่ 31 ($p = 0.490$) และข้อที่ 32 ($p = 0.467$) ซึ่งอยู่ในระดับปานกลาง

ตอนที่ 3 ผลการวิเคราะห์ค่าอำนาจจำแนกโดยภาพรวม

ค่าอำนาจจำแนกเฉลี่ยของข้อสอบทั้งหมด (41 ข้อ ไม่รวมข้อ 41) เท่ากับ 0.313 อยู่ในระดับพอใช้ถึงดี รายละเอียดการกระจายตัวแสดงในตารางที่ 3

ตารางที่ 3 สรุปการกระจายตัวของค่าอำนาจจำแนก

ระดับอำนาจจำแนก	ช่วงค่า r	จำนวนข้อ	ร้อยละ
ดีมาก	$r \geq 0.40$	10	24.4%
ดี	0.30 – 0.39	14	34.1%
พอใช้	0.20 – 0.29	10	24.4%
ต่ำ	0.01 – 0.19	6	14.6%
ลบ (ควรตัดทิ้ง)	$r < 0$	1	2.4%
รวม (ไม่รวมข้อ 41)		41	100%

จากตารางที่ 3 พบว่าข้อสอบที่มีค่าอำนาจจำแนกดีมาก ($r \geq 0.40$) มีจำนวน 10 ข้อ ได้แก่ ข้อ 1, 5, 11, 14, 19, 26, 29, 30, 35, 40 ข้อสอบที่มีค่าอำนาจจำแนกดี ($r = 0.30-0.39$) มีจำนวน 14 ข้อ ข้อสอบที่มีค่าอำนาจจำแนกพอใช้ ($r = 0.20-0.29$) มีจำนวน 10 ข้อ ข้อสอบที่มีค่าอำนาจจำแนกต่ำ ($r < 0.20$) มีจำนวน 6 ข้อ ได้แก่ ข้อ 6, 13, 18, 23, 31, 34 และข้อสอบที่มีค่าอำนาจจำแนกเป็นลบ มีจำนวน 1 ข้อ ได้แก่ ข้อ 20

ข้อสอบที่มีค่าอำนาจจำแนกสูงสุด ได้แก่ ข้อที่ 35 ($r = 0.539$) ข้อที่ 1 และ 19 ($r = 0.489$) ส่วนข้อสอบที่มีปัญหาที่สุด คือ ข้อที่ 20 ซึ่งมีค่าอำนาจจำแนกติดลบ ($r = -0.043$) ควรได้รับการแก้ไขโดยเร่งด่วน

ตอนที่ 4 ผลการวิเคราะห์ประสิทธิภาพตัวलग

การวิเคราะห์รูปแบบการตอบในแต่ละตัวเลือก (Response Pattern) พบประเด็นสำคัญดังนี้

- ข้อสอบที่มีตัวलगบางตัวไม่มีประสิทธิภาพ (มีผู้เลือกน้อยมากหรือ 0%) พบในหลายข้อ เช่น ข้อที่ 6 ที่ตัวเลือก 4 มีผู้เลือกเพียง 14% ซึ่งต่ำกว่าตัวलगอื่น บ่งชี้ว่าตัวलगบางตัวไม่สามารถดึงดูดผู้เรียนได้
- ข้อที่มีการตอบกระจายทั่วถึงทุกตัวเลือก แสดงว่าผู้เรียนไม่มีความรู้เพียงพอในเนื้อหาดังกล่าว ตัวอย่างเช่น ข้อที่ 1 ที่ตัวเลือก 3 มีผู้เลือกถึง 38% ทั้งที่คำตอบถูกต้องคือตัวเลือก 1 (21%)

3. ข้อที่ 41 ซึ่งไม่มีการกำหนดเฉลย ทำให้ไม่สามารถคำนวณค่าสถิติได้

และพบว่าการตอบกระจายเกือบเท่ากันทุกตัวเลือก (ตัวเลือก 1-4 มีผู้ตอบ 25-27%)

ซึ่งบ่งชี้ว่าผู้เรียนต้องการแนวทางที่ชัดเจน

บทที่ 5

สรุปผล อภิปรายผล และข้อเสนอแนะ

1. สรุปผลการวิจัย

การวิจัยครั้งนี้มุ่งวิเคราะห์คุณภาพข้อสอบวิชาวิทยาศาสตร์ รหัสวิชา ว21102 ระดับชั้นมัธยมศึกษาปีที่ 1 จำนวน 42 ข้อ โดยใช้ข้อมูลจากการแสกนกระดาษคำตอบด้วยโปรแกรม ZipGrade จำนวน 2,892 ฉบับ จากผู้เข้าสอบ 607 คน สรุปผลได้ดังนี้

1.1 ด้านค่าความยากง่าย

ข้อสอบทั้ง 41 ข้อ (ไม่รวมข้อ 41 ที่ไม่มีเฉลย) มีค่าความยากง่ายเฉลี่ยเท่ากับ 31.2% ($p = 0.312$) อยู่ในระดับยาก เมื่อจำแนกตามระดับ พบว่า

ข้อสอบระดับยากมาก ($p < 0.20$) จำนวน 3 ข้อ ได้แก่ ข้อที่ 4, 29, 39

ข้อสอบระดับยาก ($p = 0.20-0.40$) จำนวน 31 ข้อ ซึ่งเป็นสัดส่วนมากที่สุดถึง 76%

ข้อสอบระดับปานกลาง ($p = 0.41-0.60$) จำนวน 7 ข้อ ได้แก่ ข้อที่ 12, 21, 30, 31, 32, 33, 35

ข้อสอบระดับง่าย ($p > 0.60$) จำนวน 0 ข้อ

1.2 ด้านค่าอำนาจจำแนก

ข้อสอบทั้ง 41 ข้อ มีค่าอำนาจจำแนกเฉลี่ยเท่ากับ 0.313 อยู่ในระดับพอใช้ถึงดี โดยพบว่า

ข้อสอบที่มีค่าอำนาจจำแนกดีมาก ($r \geq 0.40$) จำนวน 10 ข้อ ได้แก่ ข้อที่ 1, 5, 11, 14, 19, 26, 29, 30, 35, 40

ข้อสอบที่มีค่าอำนาจจำแนกระดับพอใช้ถึงดี ($r = 0.20-0.39$) จำนวน 24 ข้อ

ข้อสอบที่มีค่าอำนาจจำแนกต่ำ ($r < 0.20$) จำนวน 6 ข้อ ได้แก่ ข้อที่ 6, 13, 18, 23, 31, 34 ซึ่งควรได้รับการปรับปรุง

ข้อสอบที่มีค่าอำนาจจำแนกติดลบ จำนวน 1 ข้อ ได้แก่ ข้อที่ 20 ซึ่งควรตัดทิ้งหรือแก้ไขทันที

ข้อที่ 41 ไม่สามารถคำนวณค่าอำนาจจำแนกได้ เนื่องจากไม่มีการกำหนดเฉลยในระบบ

1.3 ด้านประสิทธิภาพตัวลอง

ข้อสอบหลายข้อมีตัวลองที่ไม่มีประสิทธิภาพ โดยเฉพาะในข้อที่มีค่าอำนาจจำแนกต่ำ เช่น ข้อที่ 6, 18, 34 และข้อที่ 23 ซึ่งพบว่าตัวเลือกที่ถูกต้องมีผู้เลือกน้อยกว่าตัวลองบางตัว บ่งชี้ว่าคำถามอาจไม่ชัดเจน หรือตัวลองดูน่าเชื่อถือมากเกินไป

2. อภิปรายผล

2.1 ด้านค่าความยากง่าย

ผลการวิจัยพบว่าข้อสอบส่วนใหญ่มีค่าความยากง่ายอยู่ในระดับยาก (p เฉลี่ย = 0.312)

ซึ่งต่ำกว่าค่าที่เหมาะสม ($p = 0.50$) สอดคล้องกับผลการวิจัยของจตุพล ยงศร (2562)

ที่พบว่าข้อสอบวิทยาศาสตร์ระดับมัธยมต้นมักมีค่าความยากง่ายค่อนข้างต่ำ สาเหตุที่เป็นไปได้ ได้แก่

เนื้อหาที่ใช้ออกข้อสอบซับซ้อนเกินระดับชั้น การใช้ภาษาที่ยากเกินไปในข้อคำถาม

หรือผู้เรียนยังขาดความเข้าใจในเนื้อหาพื้นฐาน

ข้อสอบ 3 ข้อที่มีค่าความยากมากที่สุด ได้แก่ ข้อที่ 4 ($p = 0.122$) ข้อที่ 39 ($p = 0.162$) และข้อที่ 29 ($p = 0.190$) ซึ่งแม้จะยากมากแต่ยังมีค่าอำนาจจำแนกที่ยอมรับได้ ($r = 0.210, 0.295, 0.455$ ตามลำดับ) โดยเฉพาะข้อที่ 29 มีค่าอำนาจจำแนกสูงถึง 0.455 แสดงว่าแม้ข้อสอบจะยาก

แต่ยังสามารถจำแนกผู้เรียนที่เก่งและอ่อนออกจากกันได้ดี

2.2 ด้านค่าอำนาจจำแนก

โดยภาพรวมข้อสอบมีค่าอำนาจจำแนกเฉลี่ย 0.313 ซึ่งอยู่ในระดับพอใช้ มีข้อสอบที่มีอำนาจจำแนกดีมากถึง 10 ข้อ สิ่งที่น่ากังวลที่สุดคือข้อที่ 20 ซึ่งมีค่าอำนาจจำแนกติดลบ ($r = -0.043$)

หมายความว่านักเรียนกลุ่มอ่อนตอบถูกมากกว่ากลุ่มเก่ง ซึ่งบ่งชี้ว่าข้อสอบนั้นวัดสิ่งที่ไม่ต้องการวัด

อาจเกิดจากความกำกวมของคำถาม หรือตัวเลือกที่ถูกต้องไม่ชัดเจนเพียงพอ

ข้อสอบที่มีค่าอำนาจจำแนกต่ำ ($r < 0.20$) ได้แก่ ข้อที่ 6, 13, 18, 23, 31, 34

ล้วนมีปัญหาลักษณะที่แตกต่างกัน เช่น ข้อที่ 6 มีค่า $r = 0.082$ ซึ่งต่ำมาก

โดยรูปแบบการตอบพบว่าผู้เรียนเลือกตัวเลือก 1 มากที่สุด (31%) ทั้งที่คำตอบถูกต้องคือตัวเลือก 3 (26%)

แสดงว่าข้อคำถามหรือตัวเลือกไม่ชัดเจน

2.3 การเปรียบเทียบกับเกณฑ์มาตรฐาน

เมื่อพิจารณาข้อสอบที่ผ่านเกณฑ์คุณภาพทั้งสองด้าน ($p = 0.20-0.80$ และ $r \geq 0.20$)

พบว่ามีข้อสอบที่ผ่านเกณฑ์ทั้งสองด้านจำนวน 31 ข้อ คิดเป็น 75.6% ของข้อสอบทั้งหมด ซึ่งถือว่าอยู่ในระดับพอใช้

การพัฒนาคุณภาพข้อสอบในอนาคตควรมุ่งเพิ่มสัดส่วนดังกล่าวให้สูงขึ้น

3. ข้อเสนอแนะ

3.1 ข้อเสนอแนะสำหรับการปรับปรุงข้อสอบ

จากผลการวิเคราะห์ ผู้วิจัยขอเสนอแนะการจัดการข้อสอบแต่ละกลุ่มดังนี้

กลุ่มที่ 1 ข้อสอบที่ควรตัดทิ้ง

ข้อที่ 20 มีค่าอำนาจจำแนกติดลบ ควรตัดออกจากแบบทดสอบและออกข้อใหม่ทดแทน

โดยพิจารณาบททวนเนื้อหาและความชัดเจนของตัวเลือก

กลุ่มที่ 2 ข้อสอบที่ควรปรับปรุงอย่างเร่งด่วน

ข้อที่ 6, 13, 18, 23, 31, 34 มีค่าอำนาจจำแนกต่ำ ($r < 0.20$)

ควรปรับปรุงโดยทบทวนความชัดเจนของคำถาม ปรับตัวลงให้หน้าเชื่อถือ และทดสอบซ้ำก่อนนำไปใช้จริง

กลุ่มที่ 3 ข้อสอบที่ควรปรับปรุงค่าความยาก

ข้อที่ 4, 29, 39 มีค่าความยากมากเกินไป ควรปรับลดความซับซ้อน ใช้ภาษาที่เข้าใจง่ายขึ้น หรือให้ข้อมูลเพิ่มเติมในโจทย์

กลุ่มที่ 4 ข้อสอบที่มีคุณภาพดีและควรรักษาไว้

ข้อที่ 1, 5, 11, 14, 19, 26, 30, 35, 40 มีคุณภาพดีในทั้งสองด้าน

ควรเก็บไว้ในคลังข้อสอบและนำมาใช้ในการสอบครั้งต่อไป

3.2 ข้อเสนอแนะด้านการจัดการเรียนการสอน

1. ครูผู้สอนควรนำผลการวิเคราะห์ข้อสอบไปใช้ปรับปรุงกิจกรรมการเรียนการสอน

โดยเน้นเนื้อหาที่ผู้เรียนส่วนใหญ่ยังไม่เข้าใจ (จากข้อที่มี p ต่ำมาก)

2. ควรจัดกิจกรรมซ่อมเสริมสำหรับนักเรียนที่มีผลการสอบต่ำ

โดยใช้ผลการวิเคราะห์รูปแบบการตอบเพื่อระบุจุดที่นักเรียนเข้าใจผิด

3. ควรพัฒนาคลังข้อสอบที่มีคุณภาพโดยรวมรวมข้อสอบที่ผ่านการวิเคราะห์แล้วว่ามีค่าสถิติที่ดี เพื่อนำมาใช้ในการสอบครั้งต่อไป

4. ควรจัดอบรมครูในการออกข้อสอบที่มีคุณภาพ

โดยเน้นเทคนิคการเขียนตัวลงที่มีประสิทธิภาพและการกำหนดระดับพฤติกรรมการเรียนรู้

3.3 ข้อเสนอแนะสำหรับการวิจัยครั้งต่อไป

1. ควรศึกษาเปรียบเทียบผลการวิเคราะห์ข้อสอบระหว่างรายวิชา หรือระหว่างครูผู้สอนต่างคน เพื่อหาแนวทางพัฒนาข้อสอบร่วมกัน
2. ควรวิเคราะห์ความสัมพันธ์ระหว่างผลการวิเคราะห์ข้อสอบกับผลสัมฤทธิ์ทางการเรียนของนักเรียน เพื่อให้เข้าใจภาพรวมของการจัดการเรียนการสอน
3. ควรศึกษาประสิทธิภาพของโปรแกรม ZipGrade เปรียบเทียบกับโปรแกรมวิเคราะห์ข้อสอบอื่น เช่น ITEMAN หรือ R เพื่อยืนยันความน่าเชื่อถือของผล
4. ควรติดตามผลการปรับปรุงข้อสอบหลังจากนำข้อเสนอแนะไปใช้ เพื่อประเมินว่าการปรับปรุงมีประสิทธิภาพเพียงใด

4. ข้อจำกัดของการวิจัย

การวิจัยครั้งนี้มีข้อจำกัดบางประการที่ควรพิจารณา ดังนี้

1. ข้อมูลที่ใช้วิเคราะห์มาจากการแสดกนด้วยโปรแกรม ZipGrade เพียงโปรแกรมเดียว ซึ่งอาจมีข้อผิดพลาดจากการแสดกนที่ไม่สมบูรณ์ในบางฉบับ
2. ข้อที่ 41 ไม่มีการกำหนดเฉลยในระบบ ทำให้ไม่สามารถวิเคราะห์คุณภาพได้ ซึ่งอาจส่งผลต่อคะแนนรวมของผู้เรียน
3. การวิจัยครั้งนี้ไม่ได้ศึกษาความตรงเชิงเนื้อหา (Content Validity) ของข้อสอบ ซึ่งเป็นด้านคุณภาพที่สำคัญอีกด้านหนึ่ง

5. สรุป

การวิจัยครั้งนี้แสดงให้เห็นว่าแบบทดสอบวิชาวิทยาศาสตร์ รหัสวิชา ว21102 ระดับชั้นมัธยมศึกษาปีที่ 1 จำนวน 42 ข้อ มีคุณภาพในระดับพอใช้ถึงดี โดยมีค่าความยากง่ายเฉลี่ย 0.312 (ระดับยาก) และค่าอำนาจจำแนกเฉลี่ย 0.313 (ระดับพอใช้) ข้อสอบที่มีคุณภาพดีทั้งสองด้านมีจำนวน 31 ข้อ ในขณะที่มีข้อสอบที่ต้องปรับปรุงหรือตัดทิ้งรวม 7 ข้อ การใช้โปรแกรม ZipGrade พิสูจน์ให้เห็นว่าเป็นเครื่องมือที่มีประสิทธิภาพในการวิเคราะห์ข้อสอบอย่างรวดเร็ว และข้อมูลที่ได้มีประโยชน์อย่างยิ่งต่อการพัฒนาคุณภาพการวัดและประเมินผลในรายวิชาวิทยาศาสตร์

บรรณานุกรม

- จตุพล ยงศร. (2562). การวิเคราะห์คุณภาพข้อสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาวิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 1. วารสารการวัดผลการศึกษา, 36(99), 45-58.
- วิภาวดี พงศ์ไพบูลย์. (2564). ผลการใช้โปรแกรม ZipGrade ในการพัฒนาคุณภาพข้อสอบ. วารสารศึกษาศาสตร์, 32(2), 112-125.
- สมวงษ์ แปลงประสพโชค. (2551). การวัดและประเมินผลการเรียนรู้. กรุงเทพฯ: สำนักพิมพ์มหาวิทยาลัยรามคำแหง.
- อรพินท์ สอนสุภาพ. (2563). การวิเคราะห์คุณภาพข้อสอบปลายภาคเรียนด้วยโปรแกรม ZipGrade. วารสารครุศาสตร์, 48(3), 78-92.
- Bloom, B. S. (1956). Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain. New York: McKay.
- Ebel, R. L., & Frisbie, D. A. (1991). Essentials of educational measurement (5th ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Lord, F. M., & Novick, M. R. (1968). Statistical theories of mental test scores. Reading, MA: Addison-Wesley.
- Plake, B. S., Wise, S. L., & Weiss, D. J. (2019). Digital tools for classroom assessment: Efficiency and accuracy in modern testing. Educational Measurement: Issues and Practice, 38(1), 22-35.

ภาคผนวก

ตัวอย่างแบบทดสอบวัดผลปลายภาค

โรงเรียนสตรีศึกษา

ภาคเรียนที่ 2

ปีการศึกษา 2568

รายวิชาวิทยาศาสตร์ 2

รหัสวิชา ว21102

ชั้น ม.1/1-1/16

เวลา 13.00-14.30 น.

คะแนนเต็ม 30 คะแนน

ชื่อ - สกุล ชั้น.....เลขที่.....

3. หากนักเรียนนั่งเครื่องบินขึ้นไปสูงจากพื้นโลก อากาศในชั้นบรรยากาศจะมีการเปลี่ยนแปลงอย่างไร
1. ความหนาแน่นและความดันลดลง
 2. ความหนาแน่นและความดันเพิ่มขึ้น
 3. ความหนาแน่นเพิ่มขึ้นความดันลดลง
 4. ความหนาแน่นลดลงความดันเพิ่มขึ้น
6. เมื่อนักเรียนนั่งรถขึ้นบนภูเขาสูงจะรู้สึกหุ้อเนื่องจากสาเหตุใด
1. ความดันอากาศลดลง
 2. ความชื้นในอากาศมาก
 3. อุณหภูมิของอากาศเพิ่มขึ้น
 4. ปริมาณไอน้ำในอากาศมีมาก

ข้อมูลการวิเคราะห์ข้อสอบโดยใช้ โปรแกรม Zipgrade

Quiz: SciM1-42Answersheet

QUIZ DETAILS

Delete

Edit

Archive

Share

Copy

Name:	SciM1-42Answersheet
Answer Sheet:	sc21102mid66 (6558)
Date:	January 04, 2024

ANSWER KEY

Edit Answer Keys

Number Active Keys:	4
Answer Key Labels:	1234

ONLINE SUBMISSION

Question Keys Uploaded: None

Open Submission: Disabled

Number of Verified Assignments: 0

View/Edit Remote Testing Options

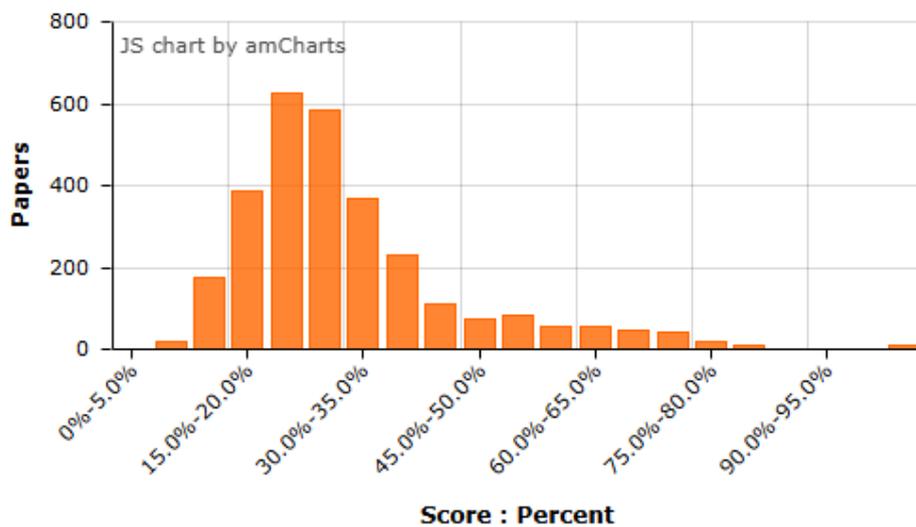
QUIZ STATISTICS

[PDF](#)[CSV](#)[Excel](#)

Number of Papers:	2893
Number of Questions:	41
Possible Points:	41

	Score	Percent
Minimum	2	4.9
Maximum	41	100.0
Average	12.8	31.2
Median	11	26.8
Std. Dev.	5.84	14.3

SCORE DISTRIBUTION



ITEM ANALYSIS

Primary Key 1 and Mapped Keys 3 - 2892 papers

This grid combines all mapped key versions relative to the primary key.

#	Pri. Ans.	# Correct	% Correct	Discrim. Factor	Responses
1	1	597.0	20.6 %	0.489	3 :38% 4 :33% 1 :21% 2 :8% _ :0% 23 :0% 14 :0% 34 :0%
2	3	804.0	27.8 %	0.380	3 :28% 1 :27% 4 :23% 2 :22% _ :0% 13 :0%
3	1	863.0	29.8 %	0.361	3 :31% 1 :30% 2 :23% 4 :16% _ :0% 23 :0% 12 :0% 14 :0% 13 :0%
4	1	353.0	12.2 %	0.210	3 :60% 2 :20% 1 :12% 4 :7% _ :0% 13 :0% 23 :0%
5	4	835.0	28.9 %	0.444	4 :29% 1 :28% 2 :23% 3 :20% _ :0%
6	3	763.0	26.4 %	0.082	1 :31% 2 :28% 3 :26% 4 :14% _ :0% 34 :0% 13 :0%
7	4	1150.0	39.8 %	0.325	4 :40% 3 :29% 2 :19% 1 :12% _ :0% 14 :0% 13 :0%
8	1	627.0	21.7 %	0.385	3 :32% 2 :27% 1 :22% 4 :20% _ :0% 23 :0%
9	3	994.0	34.4 %	0.233	2 :37% 3 :34% 1 :17% 4 :12% _ :0% 14 :0%
10	4	928.0	32.1 %	0.371	3 :32% 4 :32% 1 :19% 2 :16% _ :0% 24 :0% 34 :0% 23 :0%

11	2	1057.0	36.5 %	0.444	2 :37% 4 :35% 3 :18% 1 :11% 34 :0% 234 :0%
12	1	1173.0	40.6 %	0.328	1 :41% 3 :22% 4 :19% 2 :18% _ :0%
13	2	654.0	22.6 %	0.166	4 :35% 3 :32% 2 :23% 1 :10% _ :0% 24 :0% 13 :0% 14 :0%
14	1	1026.0	35.5 %	0.402	1 :35% 4 :32% 2 :17% 3 :15% _ :1% 34 :0%
15	1	692.0	23.9 %	0.277	3 :36% 1 :24% 2 :21% 4 :18% _ :0% 14 :0%
16	1	614.0	21.2 %	0.239	4 :37% 3 :25% 1 :21% 2 :17% _ :0% 13 :0%
17	2	710.0	24.6 %	0.375	3 :38% 2 :25% 4 :24% 1 :13% _ :0%
18	4	897.0	31.0 %	0.159	4 :31% 2 :29% 1 :22% 3 :18% _ :0% 23 :0%
19	2	1150.0	39.8 %	0.489	2 :40% 3 :23% 1 :22% 4 :15% _ :0% 23 :0% 34 :0%
20	4	1080.0	37.3 %	-0.043	4 :37% 3 :29% 2 :20% 1 :13% _ :0% 34 :0%
21	3	1309.0	45.3 %	0.340	3 :45% 2 :19% 1 :18% 4 :18% _ :0% 13 :0% 14 :0% 12 :0%
22	4	1132.0	39.1 %	0.225	4 :39% 1 :27% 2 :19% 3 :15% _ :0%

23	1	611.0	21.1 %	0.115	2:30% 4:25% 3:23% 1:21% _:0% 13:0%
24	4	594.0	20.5 %	0.238	2:33% 3:26% 1:21% 4:21% _:0%
25	2	1069.0	37.0 %	0.312	2:37% 4:28% 3:24% 1:11% 23:0% 13:0% _:0%
26	4	1145.0	39.6 %	0.448	4:40% 2:28% 3:19% 1:14% _:0%
27	2	951.0	32.9 %	0.393	2:33% 1:26% 4:21% 3:19% _:0%
28	3	886.0	30.6 %	0.331	3:31% 2:28% 4:28% 1:13% _:0% 24:0% 14:0%
29	1	550.0	19.0 %	0.455	3:36% 2:30% 1:19% 4:15% _:0% 23:0%
30	3	1255.0	43.4 %	0.462	3:43% 2:22% 4:17% 1:17% _:0% 34:0% 24:0% 13:0% 14:0%
31	3	1417.0	49.0 %	0.186	3:49% 1:31% 2:12% 4:8% _:0% 14:0% 13:0% 34:0% 23:0% 12:0%
32	1	1351.0	46.7 %	0.242	1:47% 2:19% 3:18% 4:15% _:1% 13:0% 14:0%
33	3	1172.0	40.5 %	0.369	3:41% 1:28% 4:21% 2:10% _:0% 34:0% 12:0%
34	1	766.0	26.5 %	0.063	4:35% 1:26% 3:23% 2:15% _:0% 14:0% 24:0% 13:0%
35	4	1204.0	41.6 %	0.539	4:42% 3:20% 2:20% 1:18% _:0% 14:0% 24:0% 13:0% 23:0% 12:0%

36	1	636.0	22.0 %	0.247	4:32% 3:25% 1:22% 2:21% _:0% 14:0%
37	3	754.0	26.1 %	0.383	3:26% 2:25% 4:24% 1:24% 14:0% _:0%
38	3	1126.0	38.9 %	0.254	3:39% 4:25% 2:19% 1:16% _:0% 14:0% 34:0% 23:0% 13:0%
39	1	468.0	16.2 %	0.295	2:40% 4:26% 3:17% 1:16% _:1% 34:0% 12:0%
40	2	889.0	30.7 %	0.443	2:31% 3:29% 1:21% 4:19% _:0% 23:0% 12:0% 13:0% 14:0%
41		0.0	0.0 %		4:27% 2:27% 1:25% 3:20% _:0% 13:0% 23:0% 12:0%
42	1	733.0	25.3 %	0.383	3:30% 1:25% 4:25% 2:20% _:0%