

การวิเคราะห์ข้อสอบแบบเลือกตอบหลายตัวเลือก รายวิชาสุขภาพและโรคของเด็กแรกเกิดจนถึงวัยรุ่น 1 สำหรับนิสิตแพทย์ชั้นปีที่ 4 คณะแพทยศาสตร์ มหาวิทยาลัยนเรศวร
Analysis of the Multiple Choice Questions Exam in the Course of Health and Disease of Newborns to Adolescents 1 for 4th Year Medical Students, Faculty of Medicine, Naresuan University

ภัทราวดี ยอดบุรี^{1*} อภิรัตน์ หวังธีระประเสริฐ² จิรนนท์ วีรกุล³ และ อรวรรณ กীরตสิโรจน์⁴

Patarawadee Yodburee^{1*}, Apirath Wangteeraprasert², Jiranun Weerakul³ and Orawan Keeratisiroj⁴

¹ สาขาวิชาวิทยาศาสตร์สุขภาพศึกษา คณะแพทยศาสตร์ มหาวิทยาลัยนเรศวร จังหวัดพิษณุโลก 65000

² ภาควิชาอายุรศาสตร์ คณะแพทยศาสตร์ มหาวิทยาลัยนเรศวร จังหวัดพิษณุโลก 65000

³ ภาควิชากุมารเวชศาสตร์ คณะแพทยศาสตร์ มหาวิทยาลัยนเรศวร จังหวัดพิษณุโลก 65000

⁴ สาขาวิชาอนามัยสิ่งแวดล้อม และอาชีวอนามัยและความปลอดภัย คณะสาธารณสุขศาสตร์ มหาวิทยาลัยนเรศวร จังหวัดพิษณุโลก 65000

¹ Department of Health Science Education, Faculty of Medicine, Naresuan University, Phitsanulok 65000

² Department of Medicine, Faculty of Medicine, Naresuan University, Phitsanulok 65000

³ Department of Pediatrics, Faculty of Public Health, Naresuan University, Phitsanulok 65000

⁴ Bachelor of Science Program in Environmental Health, Faculty of Public Health, Naresuan University, Phitsanulok 65000

*Corresponding author: patarawadeey@nu.ac.th

Received: December 4, 2024; Revised: February 13, 2025; Accepted: February 19, 2025

บทคัดย่อ

บทนำ การวัดประเมินผลโดยใช้การทดสอบแบบเลือกตอบหลายตัวเลือก (Multiple Choice Questions; MCQ) เป็นการวัดความรู้ได้หลายวัตถุประสงค์เนื่องจากมีจำนวนข้อมาก และหลังจากการสอบต้องมีการวิเคราะห์ข้อสอบเพื่อดูความเที่ยงและความเชื่อมั่นของข้อสอบ วัตถุประสงค์ของการศึกษานี้เพื่อวิเคราะห์คุณภาพข้อสอบ MCQs ในรายวิชาสุขภาพและโรคของเด็กแรกเกิดถึงวัยรุ่น 1 สำหรับนิสิตแพทย์ชั้นปีที่ 4 คณะแพทยศาสตร์ มหาวิทยาลัยนเรศวร โดยแบ่งเป็นการตรวจสอบคุณภาพของข้อสอบรายข้อและรายชุด รวมถึงการวิเคราะห์ความสัมพันธ์ระหว่างค่าความยาก-ง่าย ของข้อสอบก่อนและหลังสอบ

วิธีการดำเนินการวิจัย เป็นการเก็บข้อมูลย้อนหลังโดยใช้ข้อมูลจากการสอบของนิสิตแพทย์ในปีการศึกษา 2563 โดยข้อสอบมี 4 ชุด ชุดละ 100 ข้อ ชุดที่ 1 มีผู้เข้าสอบ 65 คน ชุดที่ 2 มีผู้เข้าสอบ 50 คน ชุดที่ 3 มีผู้เข้าสอบ 49 คน ชุดที่ 4 มีผู้เข้าสอบ 13 คน ก่อนการสอบมีคณะกรรมการประเมินความเที่ยงของข้อสอบให้เป็นไปตามวัตถุประสงค์ของรายวิชา หลังจากการสอบ นำผลการสอบทั้ง 4 ชุดมาวิเคราะห์ค่าความยาก-ง่าย (Item difficulty ; p) ค่าอำนาจจำแนก (Item discrimination ; r) ใช้สถิติ ค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน ค่าสัมประสิทธิ์ KR-20 ทดสอบความสัมพันธ์ โดย Pearson Correlation

ผลการศึกษา การตรวจสอบคุณภาพของข้อสอบรายข้อโดยแบ่งเป็นค่าความยาก-ง่ายของข้อสอบที่อยู่ในเกณฑ์ที่มีคุณภาพ พบว่า ข้อสอบชุดที่ 1-4 มีค่าความยากง่าย ช่วง 0.20-0.80 ร้อยละ 65.75 และค่าอำนาจจำแนกที่สามารถจำแนกผู้เข้าสอบได้ดีมากที่มีค่าตั้งแต่ 0.35 ขึ้นไป ร้อยละ 22.75 การตรวจสอบคุณภาพของข้อสอบรายชุดด้าน

ความเที่ยงตรง (Validity) เป็นไปตามวัตถุประสงค์ ตรงตามเนื้อหา สอดคล้องกับรายวิชาที่ระบุไว้ในหลักสูตร ครบถ้วน มีความตรงตามโครงสร้าง ข้อสอบสามารถวัดพฤติกรรมและสมรรถภาพด้านต่าง ๆ ของนิสิตแพทย์ได้ตามวัตถุประสงค์ที่กำหนด และเป็นไปตามหลักการของทฤษฎี โดยผ่านมติร่วมกันของคณะกรรมการรายวิชา ด้านความเชื่อมั่น (Reliability) ชุดข้อสอบที่มีค่าความเชื่อมั่น ตั้งแต่ 0.70 ขึ้นไป เป็นที่ยอมรับได้ทั้ง 4 ชุด ความสัมพันธ์ระหว่างค่าความยาก-ง่ายของข้อสอบก่อนและหลังสอบ อยู่ในช่วง 0.31–0.50 แสดงถึงความสัมพันธ์กันของข้อสอบก่อนและหลังสอบ จำนวนทั้งหมด 353 ข้อ ร้อยละ 88.25

สรุป จากการศึกษาวิเคราะห์คุณภาพข้อสอบแบบเลือกตอบแบบหลายตัวเลือก รายวิชา สุขภาพและโรคของเด็กแรกเกิดถึงวัยรุ่น 1 สำหรับนิสิตแพทย์ปี 4 พบว่าข้อสอบส่วนใหญ่มีค่าความยาก-ง่ายและค่าอำนาจจำแนกอยู่ในเกณฑ์ที่เหมาะสม ความเที่ยงตรงและความเชื่อมั่นของข้อสอบอยู่ในระดับที่ยอมรับได้ โดยมีความสัมพันธ์ระหว่างค่าความยาก-ง่ายก่อนและหลังสอบ

คำสำคัญ: การวิเคราะห์ข้อสอบ ข้อสอบ MCQs นิสิตแพทย์

Abstract

Introduction: Multiple Choice Questions (MCQs) are widely used in knowledge assessments because they can evaluate multiple objectives through numerous items. Following each exam, a test analysis is conducted to determine the validity and reliability of these questions. This study assessed the quality of MCQs used in the "Health and Diseases in Newborns to Adolescents I" course for fourth-year medical students of the Faculty of Medicine, Naresuan University. The evaluation included analyzing individual question quality and overall test set effectiveness and examining the relationship between question difficulty levels before and after the exam.

Methodology: The study was retrospective data from medical student exams in the 2020 academic year. There were four sets of 100 questions. Sets 1–4 were tested with 65, 50, 49, and 13 students. All MCQ test sets have been reviewed and assessed for quality by an examination validity assessment committee to ensure it according to the purpose of the course. The 400 answers to questions were analyzed as a difficulty (p) and discriminating (r) of the exam. Statistical tools were applied to examine relationships, including mean, standard deviation, KR–20 reliability coefficient, and Pearson correlation.

Results: The results showed that the item difficulty for all four test sets ranged from 0.20 to 0.80, with 65.75% of items falling within this range, and item discrimination values above 0.35 for 22.75% of items, indicating good discrimination power. The validity of the test sets was confirmed, as they aligned with the course objectives, content, and structure. Reliability coefficients for all four sets were found to be acceptable (≥ 0.70). The correlation between the difficulty levels of questions before and after the exam ranged from 0.31 to 0.50, suggesting consistency between pre- and post-exam difficulty levels.

Conclusion: This study analyzed the quality of multiple-choice questions (MCQs) in the Health and Diseases in Newborns to Adolescents I for 4th year medical students. The findings indicate that most questions had appropriate difficulty and discrimination indices, with validity and reliability at acceptable levels. There was a moderate correlation between pre- and post-exam difficulty levels.

Keywords: Analysis of the exam, MCQs exam, Medical students

บทนำ

การประเมินคุณภาพของแบบทดสอบ หรือเรียกอีกอย่างหนึ่งว่า การวิเคราะห์ข้อสอบ คือ การตรวจสอบคุณภาพของแบบทดสอบที่สร้างขึ้น ทั้งลักษณะที่เป็นรายข้อและรายชุด ถ้าข้อสอบข้อใดหรือข้อสอบชุดใดมีคุณภาพดีควรนำไปใช้สอบ แต่ถ้าข้อสอบมีข้อบกพร่องควรปรับปรุงแก้ไข การทำเช่นนี้จะทำให้ได้แบบทดสอบที่มีคุณภาพดีสามารถนำไปทดสอบ ช่วยให้การวัดและประเมินผลมีประสิทธิภาพดียิ่งขึ้น ทั้งยังสามารถนำข้อสอบที่มีคุณภาพเก็บไว้เป็นคลังข้อสอบเพื่อใช้งานต่อไปในอนาคต และยังเป็นการใช้วิธีการทางสถิติเพื่อวิเคราะห์คำตอบที่ผู้สอบตอบข้อสอบในการสอบ เพื่อประเมินว่าข้อสอบที่นำมาใช้ในการสอบครั้งนั้น มีคุณสมบัติของข้อสอบที่สามารถวัดผลได้ตรงตามที่กำหนดวัตถุประสงค์ของรายวิชา (เชิดศักดิ์ ไอรมณีรัตน์, 2552) การพิจารณาความเหมาะสมและความถูกต้องของข้อสอบที่สร้างขึ้น โดยมีวัตถุประสงค์เพื่อต้องการตรวจสอบคุณสมบัติและพัฒนาข้อสอบให้เป็นข้อสอบที่ดีตามเกณฑ์ที่กำหนดไว้ ซึ่งการวิเคราะห์ข้อสอบแบบเลือกตอบ (multiple choice question ; MCQ) จะให้ผลทางสถิติที่จะนำมาใช้วิเคราะห์คุณสมบัติและคุณภาพของข้อสอบได้ ได้แก่ ค่าความยาก-ง่ายของข้อสอบ (item difficulty ; p) และค่าอำนาจจำแนก (item discrimination ; r) ของข้อสอบที่ได้จากการวิเคราะห์ผลหลังการสอบ (ธนภรณ์ แดงใจ และคณะ, 2561)

ภาควิชากุมารเวชศาสตร์ คณะแพทยศาสตร์ มหาวิทยาลัยนเรศวร ได้มีนโยบายในการจัดการเรียน การสอนเกี่ยวกับเรื่องการวัดและประเมินผล ให้มีมาตรฐานเดียวกัน เนื่องจากคณะแพทยศาสตร์ ได้มีการจัดการเรียนการสอนร่วมกันกับเครือข่าย ศูนย์แพทยศาสตรศึกษา ชั้นคลินิก ทั้ง 5 ศูนย์แพทย์ (รพ.พุทธชินราช พิษณุโลก, รพ.อุตรดิตถ์, รพ.แพร่, รพ.พิจิตร และ รพ.สมเด็จพระเจ้าตากสินมหาราช) ซึ่งในรายวิชาสุขภาพและโรคของเด็กแรกเกิดจนถึงวัยรุ่น 1 จึงได้สนองนโยบายของคณะแพทยศาสตร์ ในเรื่องการจัดทำคลังข้อสอบ เพื่อใช้ร่วมกันกับเครือข่าย โดยได้นำข้อสอบแบบเลือกตอบหลายตัวเลือก จำนวน 400 ข้อ ที่ออกโดยยึดเกณฑ์แพทย์สภาเป็นกรอบในการออกข้อสอบจัดทำตารางกำหนดสัดส่วนข้อสอบ (Table of Specification) และขอความร่วมมือจากอาจารย์ทุกท่านของทุกศูนย์แพทย์ในการออกข้อสอบตามหัวข้อที่กำหนด มีตัวเลือก 5 ตัวเลือก การหาคุณภาพจะมีการหาคุณภาพด้านความเที่ยงตรง โดยคณะกรรมการพิจารณาข้อสอบของรายวิชา มาเข้าร่วมการประชุมพัฒนาปรับปรุงข้อสอบเพื่อพิจารณาความตรงด้านเนื้อหาและความเหมาะสมของข้อคำถาม ตัวถูก ตัวลวง เพื่อคัดเลือกข้อสอบที่ผ่านการพัฒนาปรับปรุงแล้วมาจัดเป็นชุดข้อสอบ จำนวน 4 ชุด ชุดละ 100 ข้อ แล้วนำมาใช้สอบวัดความรู้ความสามารถของนิสิตแพทย์ชั้นปีที่ 4 ที่ขึ้นเรียนในรายวิชา สุขภาพและโรคของเด็กแรกเกิดจนถึงวัยรุ่น 1 ของปีการศึกษา 2563 ซึ่งยังขาดการหาคุณภาพรายข้อเกี่ยวกับค่าความยาก-ง่าย ค่าอำนาจจำแนก และการหาคุณภาพข้อสอบรายชุดเกี่ยวกับความเที่ยงตรงและความเชื่อมั่นของข้อสอบ

การวิเคราะห์ข้อสอบตามทฤษฎีการสอบแบบดั้งเดิม เพื่อทดสอบคุณภาพของข้อสอบประกอบไปด้วย 2 ส่วน คือ 1) การวิเคราะห์ข้อสอบรายข้อ ต้องพิจารณาถึงความยาก-ง่ายของข้อสอบ (item difficulty ; p) ความสามารถในการจำแนกผู้สอบตามระดับความสามารถ (item discrimination ; r) และประสิทธิภาพของตัวลวง (distractor functionality) 2) การวิเคราะห์ข้อสอบรายชุด (test analysis) เป็นการพิจารณาถึงความเที่ยงตรงของคะแนนสอบ (internal consistency reliability) การกระจายตัวของคะแนน และคะแนนเฉลี่ย (standard deviation and mean score) ค่าความยากง่ายเฉลี่ยของข้อสอบ (average difficulty) และค่าความสามารถในการแยกแยะผู้สอบเฉลี่ย (average discrimination) (เชิดศักดิ์ ไอรมณีรัตน์, 2552)

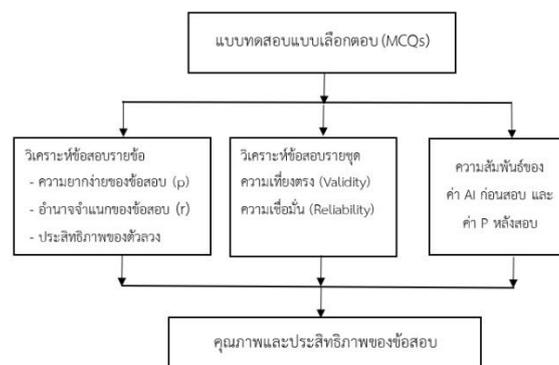
ข้อสอบแบบปรนัย หรือ แบบทดสอบแบบเลือกตอบหลายตัวเลือก ประโยชน์ในการประเมินประสิทธิภาพของผู้เรียน มีลักษณะที่เป็นกลาง ตัดปัญหาความเอนเอียงของผู้ตรวจข้อสอบ มีความน่าเชื่อถือและความถูกต้อง การวิเคราะห์ข้อสอบจึงเป็นเครื่องมือวิเคราะห์คุณภาพข้อสอบเพื่อปรับปรุงข้อบกพร่องของข้อสอบให้เหมาะสม และนำกลับไปใช้ทดสอบอีกครั้ง หรือจัดการปรับปรุงข้อสอบที่ไม่เหมาะสม การหาความสัมพันธ์ระหว่างค่าความยาก-ง่าย ค่าอำนาจการจำแนก และประสิทธิภาพตัวลวง ของข้อสอบทั้งแบบรายข้อ และรายชุด จะช่วยให้สามารถสังเกตลักษณะข้อสอบและปรับปรุงคุณภาพของการสอบ การแก้ไขปรับปรุงข้อสอบช่วยให้ระบุข้อสอบที่ยากเกินไป หรือ ง่ายเกินไป ข้อสอบที่ไม่สามารถแยกความแตกต่างระหว่างผู้เรียนได้

ผู้วิจัยเป็นนักวิชาการศึกษาคณะครุศาสตร์ และปฏิบัติงานเกี่ยวกับการจัดการเรียนการสอนในภาควิชาศึกษาศาสตร์ คณะแพทยศาสตร์ มหาวิทยาลัยนเรศวร จึงมีความสนใจที่จะทำการวิเคราะห์ข้อสอบ หาคุณภาพของข้อสอบทั้งรายข้อและรายชุด เพื่อสามารถนำผลการวิเคราะห์ข้อสอบที่ได้จากการศึกษาไปพัฒนาปรับปรุงข้อสอบให้มีคุณภาพและมาตรฐาน สามารถนำข้อสอบเก็บเข้าคลังข้อสอบของรายวิชาต่อไป

วัตถุประสงค์

1. วิเคราะห์คุณภาพข้อสอบรายข้อ ความยาก-ง่ายของข้อสอบ (item difficulty ; p) ความสามารถในการจำแนกผู้สอบตามระดับความสามารถ (item discrimination ; r) และประสิทธิภาพของตัวลวง (distractor functionality)
2. วิเคราะห์คุณภาพข้อสอบรายชุด ความเที่ยงตรง (validity) และความเชื่อมั่น (reliability)
3. วิเคราะห์ความสัมพันธ์ระหว่างค่าความยากง่ายของข้อสอบ (acceptability index ; AI) ก่อนสอบ กับ ค่าความยากง่ายของข้อสอบ (item difficulty ; p) หลังการสอบ

กรอบแนวคิด



ภาพที่ 1 กรอบแนวคิดในการวิจัย

วิธีการศึกษา

การวิจัยครั้งนี้ ผู้วิจัยขอเสนอวิธีการดำเนินการวิจัย ตามหัวข้อดังต่อไปนี้

1. ประชากร

ข้อมูลคะแนนสอบข้อสอบปรนัย ของนิสิตแพทย์ชั้นปีที่ 4 ที่ขึ้นเรียนรายวิชา สุขภาพและโรคของเด็กแรกเกิด จนถึงวัยรุ่น 1 ปีการศึกษา 2563 ทุกศูนย์แพทย์ จำนวน 177 คน

2. กลุ่มตัวอย่าง

ข้อมูลที่ใช้ในการวิจัยครั้งนี้ เป็นข้อมูลคะแนนสอบของนิสิตแพทย์ชั้นปีที่ 4 ที่ขึ้นเรียนรายวิชา สุขภาพและโรคของเด็กแรกเกิดจนถึงวัยรุ่น 1 ปีการศึกษา 2563 ทั้ง 6 ศูนย์แพทย์ รวมจำนวน 177 คน แบ่งตามศูนย์แพทย์ คณะแพทยศาสตร์ รพ.มหาวิทยาลัยนเรศวร มีจำนวนนิสิต 34 คน แบ่งเป็น 3 กลุ่มๆ ละ 11-12 คน ศูนย์แพทย์ ศึกษาศาสตร์ รพ.อุตรดิตถ์ มีจำนวนนิสิต 27 คน แบ่งเป็น 3 กลุ่มๆ ละ 9 คน ศูนย์แพทย์ศึกษาศาสตร์ รพ.แพร่ มีจำนวนนิสิต 18 คน แบ่งเป็น 3 กลุ่มๆ ละ 6 คน ศูนย์แพทย์ศึกษาศาสตร์ รพ.พิจิตร มีจำนวนนิสิต 21 คน แบ่งเป็น 3 กลุ่มๆ ละ 7 คน ศูนย์แพทย์ศึกษาศาสตร์ รพ.สมเด็จพระเจ้าตากสินมหาราช มีจำนวนนิสิต 15 คน แบ่งเป็น 3 กลุ่มๆ ละ 5 คน และศูนย์แพทย์ศึกษาศาสตร์ รพ.พุทธชินราช พิษณุโลก มีจำนวนนิสิต 62 คน แบ่งเป็น 5 กลุ่มๆ ละ 12-13 คน

ในการศึกษาวิจัยครั้งนี้ เป็นการศึกษาค้นคว้าข้อมูลการตอบข้อสอบรายวิชา สุขภาพและโรคของเด็กแรกเกิดจนถึงวัยรุ่น 1 ปีการศึกษา 2563 ของนิสิตแพทย์ชั้นปีที่ 4 ซึ่งได้มีการกำหนดการเลือกชุดข้อสอบตามกลุ่มที่ขึ้นเรียนในช่วงเดียวกันให้ใช้ข้อสอบชุดเดียวกัน และมีการกำหนดให้มีการจัดสอบในวัน และเวลาเดียวกัน โดยผู้วิจัยได้ใช้เทคนิค ร้อยละ 27 ของ จุง เดห์ ฟาน มาหาค่าความยาก-ง่าย และค่าอำนาจการจำแนกซึ่งเหมาะสมนำมาใช้กับข้อสอบที่มีผู้เข้าสอบตั้งแต่ 30 คนขึ้นไป (จักรพงษ์ แผ่นทอง, 2564) โดยนำผลการตรวจให้คะแนนผู้เข้าสอบมาจัดเรียงตามลำดับจากผู้ที่ได้คะแนนสูงไปหาคะแนนต่ำ แบ่งกลุ่มคะแนน เป็นกลุ่มคะแนนสูง และกลุ่มคะแนนต่ำ ร้อยละ 27 นำไปคำนวณหาค่าทางสถิติต่าง ๆ ที่กำหนดไว้ในงานวิจัย

3. เครื่องมือที่ใช้ในการวิจัย

ผลการตอบข้อสอบแบบเลือกตอบหลายตัวเลือกของ รายวิชา สุขภาพและโรคของเด็กแรกเกิดจนถึงวัยรุ่น 1 ที่ใช้ทดสอบกับนิสิตแพทย์ชั้นปีที่ 4 ที่ขึ้นเรียนของคณะแพทยศาสตร์ มหาวิทยาลัยนเรศวร โดยมีจำนวนข้อสอบทั้งหมด 400 ข้อ นำมาจัดเป็นชุดข้อสอบ 4 ชุด ชุดละ 100 ข้อ เลือกใช้ชุดข้อสอบ ด้วยวิธีการสุ่ม โดยพิจารณาค่า p ใกล้เคียงกัน มีการกำหนดสัดส่วนของการวัดความรู้ในด้านต่าง ๆ ตามวัตถุประสงค์รายวิชาที่กำหนดไว้ในทุกชุดเท่า ๆ กัน คือ มีข้อสอบที่วัดเรื่องการตรวจประเมิน การวินิจฉัย การรักษาในจำนวนที่ใกล้เคียงกัน และกระจายไปในข้อสอบทั้ง 4 ชุด

4. วิธีดำเนินการในการเก็บรวบรวมข้อมูล

การศึกษาในครั้งนี้ เป็นการนำข้อมูลแบบทฤษฎีภูมิ นำผลคะแนนสอบ ผลการตอบของผู้เข้าสอบของข้อสอบรายวิชาสุขภาพและโรคของเด็กแรกเกิดจนถึงวัยรุ่น 1 ของนิสิตแพทย์ชั้นปีที่ 4 ที่ขึ้นเรียนเมื่อปีการศึกษา 2563 โดยมีวิธีดำเนินการในการเก็บรวบรวมข้อมูล ดังนี้

สิ่งที่ได้ดำเนินการแล้ว ก่อนนำข้อมูลมาทำการศึกษา

1. จัดทำตารางกำหนดสัดส่วนข้อสอบ ของรายวิชาสุขภาพและโรคของเด็กแรกเกิดจนถึงวัยรุ่น 1 เพื่อให้ อาจารย์แต่ละศูนย์แพทย์จำนวน 6 ศูนย์ ออกข้อสอบทั้งหมด 400 ข้อ

2. นำข้อสอบทั้งหมดมาประชุมพิจารณาถ้อยแถลงข้อสอบ พร้อมกำหนดค่า AI ก่อนการสอบร่วมกันกับศูนย์แพทย์เครือข่าย โดยคณะกรรมการรายวิชาสุขภาพและโรคของเด็กแรกเกิดจนถึงวัยรุ่น 1

3. นำข้อสอบที่ได้จากการถ้อยแถลงข้อสอบร่วมกันกับศูนย์แพทย์ จำนวน 400 ข้อ มาจัดชุดข้อสอบจำนวน 4 ชุด ชุดละ 100 ข้อ

4. เมื่อถึงกำหนดวันสอบลงกอง ได้นำแบบทดสอบชุดที่ได้เลือกไว้ นำไปทดสอบกับนิสิตแพทย์ของแต่ละศูนย์แพทย์ ในวันและเวลาเดียวกัน

5. นำข้อสอบที่ผ่านการทดสอบแล้ว มาตรวจวิเคราะห์ข้อสอบด้วยโปรแกรมตรวจวิเคราะห์ข้อสอบของคณะแพทยศาสตร์ มหาวิทยาลัยนเรศวร ตามดัชนีค่าอำนาจจำแนกโดยวิธี ค่าความยากจากตารางจุง-เตห์-ฟาน (Chung-The-Fan) ของคณะแพทยศาสตร์ ศิริราชพยาบาล

6. ดำเนินการขอจริยธรรมการวิจัยในมนุษย์ จากคณะกรรมการจริยธรรมการวิจัยในมนุษย์ กลุ่มสาขาวิทยาศาสตร์ทางการแพทย์ มหาวิทยาลัยนเรศวร โดยผลพิจารณาโครงการได้รับการรับรอง แบบยกเว้น หมายเลขโครงการ P3-0002-2566 ลงวันที่ 20 มกราคม 2566

สิ่งที่จะนำมาศึกษาวิจัยในครั้งนี้

นำข้อมูลการตอบของผู้เข้าสอบรายวิชาสุขภาพและโรคของเด็กแรกเกิดจนถึงวัยรุ่น 1 ปีการศึกษา 2563 ที่ได้จากการตรวจด้วยเครื่องตรวจวิเคราะห์ข้อสอบ มากำหนดเป็นตัวเลข 1 (คำตอบที่ถูกต้อง) และ 0 (คำตอบที่ผิด) เพื่อนำไปวิเคราะห์หาค่าความยาก-ง่ายของข้อสอบ ค่าอำนาจจำแนกของข้อสอบ และค่าความเชื่อมั่นของชุดข้อสอบ โดยใช้โปรแกรม Microsoft Excel 2010 และโปรแกรม SPSS เวอร์ชัน 26.0 เพื่อหาค่าที่กำหนดไว้ในงานวิจัยต่อไป

5. สถิติที่ใช้ในการวิเคราะห์ข้อมูล

ค่าเฉลี่ย ร้อยละ ส่วนเบี่ยงเบนมาตรฐาน

5.1 วิเคราะห์ข้อสอบรายข้อ

- ค่าความยาก-ง่ายของข้อสอบ (difficulty index, p) คือ นำผลการตอบข้อสอบของนิสิตแพทย์มาคำนวณทางคณิตศาสตร์ เพื่อหาสัดส่วนของจำนวนผู้ที่ตอบข้อสอบได้ถูกต้องต่อจำนวนผู้ที่ตอบข้อสอบทั้งหมดของข้อสอบโดยใช้โปรแกรม Microsoft Excel 2010 ซึ่งค่า p ระหว่าง 0.20 – 0.80 ถือว่าเป็นข้อสอบที่ดี มีความยากพอเหมาะ (Nunnally, 1967)

- อำนาจจำแนกของข้อสอบ (discrimination index, r) นำผลการตอบข้อสอบของนิสิตแพทย์มาดำเนินการคำนวณทางคณิตศาสตร์ เพื่อดูความสามารถของข้อสอบในการจำแนกหรือแยกกลุ่มผู้สอบเป็น 2 กลุ่ม คือ กลุ่มเก่ง และกลุ่มอ่อน โดยใช้โปรแกรม Microsoft Excel 2010 ค่าที่ได้มากกว่าหรือเท่ากับ 0.35 แสดงถึงเป็นข้อสอบดีมาก มีความสามารถในการจำแนกสูง ค่าระหว่าง 0.25–0.35 เป็นที่ยอมรับได้ ค่าน้อยกว่า 0.25 ต้องปรับปรุงข้อสอบ (จรอ ประดับศรี, 2550)

- ประสิทธิภาพตัวลวง (distractor effectiveness) นำผลการตอบข้อสอบของนิสิตแพทย์มาคำนวณทางคณิตศาสตร์ เพื่อหาสัดส่วนของจำนวนคนที่เลือกตอบตัวลวง โดยใช้โปรแกรม Microsoft Excel 2010 ตัวลวงที่ดีควรมีค่า Pw ตั้งแต่ 0.05 ขึ้นไป (จรอ ประดับศรี, 2550)

5.2 วิเคราะห์ข้อสอบรายชุด

- ความเที่ยงตรง (validity) ความตรงตามโครงสร้าง (construct Validity) คือ ผู้วิจัยได้อยู่ในกระบวนการจัดทำข้อสอบของรายวิชาผ่านการพิจารณาข้อสอบที่ละข้อจากคณะกรรมการรายวิชา โดยมีการจัดทำข้อสอบให้มีความสอดคล้องกับเนื้อหาของรายวิชา เป็นไปตามวัตถุประสงค์ที่กำหนดไว้ในรายวิชา และเป็นไปตามหลักการของทฤษฎีสามารถวัดพฤติกรรมและสมรรถนะด้านต่าง ๆ ได้ตามวัตถุประสงค์ที่กำหนดไว้ โดยมีการจัดทำตารางกำหนดสัดส่วน

ข้อสอบ สร้างข้อสอบให้สอดคล้องตรงตามเนื้อหาวิชาที่ระบุไว้ในรายวิชา อย่างครบถ้วน มีการตรวจสอบความถูกต้องของข้อสอบโดยผ่านการพิจารณาจากคณะกรรมการรายวิชาทั้ง 6 ศูนย์แพทย์

- ความเชื่อมั่น (reliability) นำผลการตอบข้อสอบของนิสิตแพทย์มากำหนดเป็นตัวเลข 1 และ 0 โดยใช้โปรแกรม Microsoft Excel 2010 และหาค่าสัมประสิทธิ์ KR-20 โดยใช้โปรแกรม SPSS เวอร์ชัน 26.0 เพื่อหาค่าความเชื่อมั่นของชุดข้อสอบ ค่า KR-20 ที่มีค่ามากกว่าหรือเท่ากับ 0.7 ขึ้นไป ถือเป็นชุดข้อสอบเป็นที่ยอมรับได้ (พวงพะยอม แก้วพิลา และคณะ, 2549)

5.3 ทดสอบความสัมพันธ์

นำค่าความยาก-ง่ายของข้อสอบก่อนการสอบ (AI) และค่าความยาก-ง่ายของข้อสอบหลังการสอบ (p) มาทดสอบความสัมพันธ์ทางสถิติ Pearson Correlation ผ่านโปรแกรม SPSS เวอร์ชัน 26.0 ค่าที่มากกว่าหรือเท่ากับ 0.7 ขึ้นไป ถือว่ามีความสัมพันธ์กันในระดับสูง (ธนาภรณ์ แดงใจ และคณะ, 2561)

ผลการศึกษา

ในการศึกษาครั้งนี้ เป็นการศึกษาถึงคุณภาพของแบบทดสอบแบบหลายตัวเลือก โดยมีวัตถุประสงค์การศึกษาเพื่อทดสอบคุณภาพข้อสอบรายข้อ และรายชุด รวมถึงทดสอบความสัมพันธ์กันของค่า AI และค่า p โดยใช้ทฤษฎีและสถิติที่กำหนดไว้ในการศึกษา วิเคราะห์ข้อมูลผลการตอบข้อสอบจากผู้สอบข้อสอบ รายวิชาสุขภาพและโรคของเด็กแรกเกิดจนถึงวัยรุ่น 1 สำหรับนิสิตแพทย์ชั้นปีที่ 4 คณะแพทยศาสตร์ มหาวิทยาลัยนเรศวร ที่ขึ้นเรียนในปีการศึกษา 2563 โดยผู้วิจัยขอเสนอผลการศึกษาในหัวข้อตามลำดับดังนี้

1. ผลการวิเคราะห์คุณภาพข้อสอบรายข้อ ข้อสอบที่มีค่าความยาก-ง่ายของข้อสอบ ระหว่าง 0.20 – 0.80 มีจำนวน 263 ข้อ คิดเป็นร้อยละ 65.75 ข้อสอบที่มีค่าความสามารถในการจำแนกผู้สอบตามระดับความสามารถ ที่มีค่าตั้งแต่ 0.35 ขึ้นไป มีจำนวน 91 ข้อ คิดเป็นร้อยละ 22.75 และประสิทธิภาพของตัวลองที่ดีมีค่าตั้งแต่ 0.05 ขึ้นไป มีจำนวน 274 ข้อ คิดเป็นร้อยละ 68.50

2. ผลการวิเคราะห์คุณภาพข้อสอบรายชุด ได้แก่ ความเที่ยงตรง และความเชื่อมั่น ข้อสอบทั้ง 4 ชุด มีความเที่ยงตรง เป็นไปตามวัตถุประสงค์ที่กำหนดไว้ในรายวิชาและเป็นไปตามหลักการของทฤษฎี สามารถวัดพฤติกรรมและสมรรถภาพด้านต่าง ๆ ได้ตามวัตถุประสงค์ที่กำหนดไว้ มีค่า KR-20 ชุดที่ 1 = 0.779 ชุดที่ 2 = 0.789 ชุดที่ 3 = 0.821 ชุดที่ 4 = 0.821 ซึ่งมีค่ามากกว่าหรือเท่ากับ 0.7 ขึ้นไป ถือเป็นชุดข้อสอบเป็นที่ยอมรับได้

3. ผลการทดสอบความสัมพันธ์ระหว่างค่าความยากง่ายของข้อสอบก่อนการสอบและค่าความยากง่ายของข้อสอบหลังการสอบมีความสัมพันธ์กันจำนวน 353 ข้อ คิดเป็นร้อยละ 88.25

ความยาก-ง่ายของข้อสอบ (p) หลังการสอบ พบว่ามีข้อสอบที่มีความยากพอเหมาะ (ดีมาก) จำนวน 80 ข้อ คิดเป็นร้อยละ 20 พบข้อสอบที่ง่ายมาก จำนวน 106 ข้อ คิดเป็นร้อยละ 26.50 พบข้อสอบที่ยากมาก จำนวน 31 ข้อ คิดเป็นร้อยละ 7.75 จากข้อสอบทั้งหมด (ตารางที่ 1)

ตารางที่ 1 ตารางสรุประดับความยาก-ง่ายของข้อสอบ (p) หลังการสอบ

ระดับความยากง่าย ของข้อสอบ (p)	จำนวน ข้อ	ค่า p หลังการสอบ					
		ชุดที่ 1	ชุดที่ 2	ชุดที่ 3	ชุดที่ 4	รวม	ร้อยละ
ง่ายมาก 0.81-1.00	-	24	23	25	34	106	26.50
ค่อนข้างง่าย (ดี) 0.60-0.80	10-15	29	22	29	20	100	25.00
ยากพอเหมาะ (ดีมาก) 0.40-0.59	60-90	26	20	21	13	80	20.00
ค่อนข้างยาก (ดี) 0.20-0.39	10-15	16	30	19	18	83	20.75
ยากมาก 0-0.19	-	5	5	6	15	31	7.75

หมายเหตุ: ในแต่ละชุดข้อสอบมีจำนวน 100 ข้อ

อำนาจการจำแนกของข้อสอบ พบว่ามีข้อสอบที่ดีมาก มีความสามารถในการจำแนกสูงสามารถแยกผู้สอบได้ดีอยู่ในชุดที่ 2 และ 3 คือ 30 และ 26 ข้อตามลำดับ รวมข้อสอบทั้ง 4 ชุด มีทั้งหมด 91 ข้อ คิดเป็นร้อยละ 22.75 และพบข้อสอบที่ต้องปรับปรุงถึง 236 ข้อ คิดเป็นร้อยละ 59 จากข้อสอบทั้งหมด (ตารางที่ 2)

ตารางที่ 1 ตารางสรุปอำนาจการจำแนกของข้อสอบทั้ง 4 ชุด

อำนาจจำแนกของข้อสอบ (r)	ผลการวิเคราะห์					
	ชุดที่ 1	ชุดที่ 2	ชุดที่ 3	ชุดที่ 4	รวม	ร้อยละ
ข้อสอบดีมาก มีความสามารถ ในการจำแนกสูง ≥ 0.35	14	30	26	21	91	22.75
ข้อสอบเป็นที่ยอมรับได้ 0.25-0.35	25	10	10	28	73	18.25
ต้องปรับปรุงข้อสอบ < 0.25	61	60	64	51	236	59.00

หมายเหตุ: ในแต่ละชุดข้อสอบมีจำนวน 100 ข้อ

จากการเปรียบเทียบผลการวิเคราะห์ข้อสอบทั้ง 4 ชุด พบว่า ข้อสอบที่มีค่า p น้อยกว่า 0.20 มีจำนวน 31 ข้อ คิดเป็นร้อยละ 7.75 ข้อสอบที่มีค่า p มากกว่า 0.80 จำนวน 106 ข้อ คิดเป็นร้อยละ 26.50 ข้อสอบที่มีค่า r น้อยกว่า 0.25 จำนวน 236 ข้อ คิดเป็นร้อยละ 59 ข้อสอบที่มีค่า AI ก่อนการสอบ สัมพันธ์กับค่า P หลังการสอบ จำนวน 353 ข้อ คิดเป็นร้อยละ 88.25 ประสิทธิภาพของตัวลง ที่มีค่าตั้งแต่ 0.05 จำนวน 274 ข้อ คิดเป็นร้อยละ 68.5 แสดงให้เห็นถึงคุณภาพข้อสอบที่มีความเหมาะสม (ตารางที่ 3)

ตารางที่ 2 ตารางสรุปเปรียบเทียบผลการวิเคราะห์ข้อสอบทั้ง 4 ชุด

ผลการวิเคราะห์	ชุดที่ 1	ชุดที่ 2	ชุดที่ 3	ชุดที่ 4	รวม	ร้อยละ
จำนวนนิสิตแพทย์ (คน)	65	50	49	13	177	100.00
จำนวนข้อสอบ (ข้อ)	100	100	100	100	400	100.00
จำนวนข้อสอบที่มีค่า P น้อยกว่า 0.20 (ข้อ)	5	5	6	15	31	7.75
จำนวนข้อสอบที่มีค่า P มากกว่า 0.80 (ข้อ)	24	23	25	34	106	26.50

ผลการวิเคราะห์	ชุดที่ 1	ชุดที่ 2	ชุดที่ 3	ชุดที่ 4	รวม	ร้อยละ
จำนวนข้อสอบที่มีค่า r น้อยกว่า 0.25 (ข้อ)	61	60	64	51	236	59.00
ข้อสอบที่มีค่า AI สัมพันธ์กับค่า P (ข้อ)	90	89	88	86	353	88.25
ค่า R _w ตั้งแต่ 0.05 (ข้อ)	68	63	75	68	274	68.50

หมายเหตุ: เนื่องจากข้อสอบชุดที่ 4 มีจำนวนผู้เข้าสอบเพียง 13 คน ซึ่งน้อยกว่า 30 คน จึงไม่เหมาะแก่การนำมาวิเคราะห์

ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่า AI ของข้อสอบที่กำหนดไว้ก่อนการสอบ และค่า P ที่ได้หลังการสอบเสร็จของข้อสอบชุด 1-2 มีค่าที่ใกล้ 0.50 ถือเป็นความสัมพันธ์ปานกลาง และชุดที่ 3-4 มีค่าที่ต่ำกว่า 0.30 ถือเป็นความสัมพันธ์ที่ต่ำ ค่า P-value มีความสัมพันธ์กันในเชิงบวกอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ในข้อสอบชุดที่ 1-3 ซึ่งเหมาะสมสำหรับเนื้อหาที่ควรรู้ ในขณะที่ชุดที่ 4 ไม่มีความสัมพันธ์ เนื่องจากจำนวนผู้เข้าสอบน้อย (ตารางที่ 4)

ตารางที่ 3 ตารางแสดงค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สันของข้อสอบทั้ง 4 ชุด

ข้อสอบ	ผลการวิเคราะห์	
	ค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน	P-value
ชุดที่ 1	0.310	.002
ชุดที่ 2	0.315	.001
ชุดที่ 3	0.294	.003
ชุดที่ 4	0.165	.101

อภิปรายและสรุปผล

จากการศึกษาวิจัยเรื่อง การวิเคราะห์ข้อสอบ MCQ รายวิชาสุขภาพและโรคของเด็กแรกเกิดจนถึงวัยรุ่น 1 สำหรับนิสิตแพทย์ชั้นปีที่ 4 คณะแพทยศาสตร์ มหาวิทยาลัยนเรศวร ได้อภิปรายผลการวิจัยดังนี้

จากวัตถุประสงค์การวิจัย เพื่อศึกษาวิเคราะห์คุณภาพข้อสอบรายข้อ เรื่อง ความยาก-ง่ายของข้อสอบ ความสามารถในการจำแนกผู้สอบตามระดับความสามารถ และประสิทธิภาพของตัวลวง การวิเคราะห์คุณภาพข้อสอบรายชุด ความเที่ยงตรง และความเชื่อมั่น และเปรียบเทียบค่าความยากง่ายของข้อสอบรายข้อ ระหว่างค่าความยาก-ง่ายของข้อสอบก่อนสอบกับค่าความยาก-ง่ายของข้อสอบหลังการสอบของข้อสอบ MCQ รายวิชาสุขภาพและโรคของเด็กแรกเกิดจนถึงวัยรุ่น 1 สำหรับนิสิตแพทย์ชั้นปีที่ 4 จำนวน 177 คน ปีการศึกษา 2563 มีข้อสอบทั้งหมด 4 ชุด ชุดละ 100 ข้อ แต่ละข้อมีตัวเลือก 5 ตัวเลือก โดยใช้สถิติเชิงบรรยาย ด้วยโปรแกรม SPSS พบว่า ข้อสอบทั้ง 4 ชุด เป็นข้อสอบที่มีค่าความยาก-ง่ายของข้อสอบที่มีค่าตั้งแต่ 0.20 – 0.80 จำนวน 263 ข้อ คิดเป็นร้อยละ 65.75 อำนาจการจำแนกของข้อสอบ พบข้อสอบที่ดีมาก มีความสามารถในการจำแนกสูงสามารถแยกผู้สอบได้ดี มีทั้งหมด 91 ข้อ คิดเป็นร้อยละ 22.75 และพบข้อสอบที่ต้องปรับปรุงถึง 236 ข้อ คิดเป็นร้อยละ 59 จากข้อสอบทั้งหมด ซึ่งเป็นข้อสอบที่มีค่าอำนาจจำแนกของข้อสอบในระดับต่ำ และประสิทธิภาพของตัวลวงที่ดี มีค่าตั้งแต่ 0.05 ขึ้นไป มีจำนวน 274 ข้อ คิดเป็นร้อยละ 68.50 สอดคล้องกับการศึกษาของ Kennedy ที่ศึกษาถึงดัชนีความยาก-ง่าย และอำนาจการจำแนกของข้อสอบ พบว่า

ข้อสอบมีค่าดัชนีความยาก-ง่ายของข้อสอบ $p = 40-60$ ซึ่งอยู่ในระดับความยาก-ง่ายที่เหมาะสม และค่าอำนาจการจำแนก มีค่าเฉลี่ย 0.22 อยู่ในระดับดีพอสมควร

การวิเคราะห์คุณภาพข้อสอบรายชุด ความเที่ยงตรง และความเชื่อมั่น ข้อสอบทั้ง 4 ชุด มีความเที่ยงตรง เป็นไปตามวัตถุประสงค์ที่กำหนดไว้ในรายวิชา และเป็นไปตามหลักการของทฤษฎี สามารถวัดพฤติกรรมและสมรรถนะด้านต่าง ๆ ได้ตามวัตถุประสงค์ที่กำหนดไว้ มีค่า KR-20 ที่มีค่ามากกว่าหรือเท่ากับ 0.70 ขึ้นไป ถือเป็นชุดข้อสอบที่ยอมรับได้ ไม่สอดคล้องกับการศึกษาของพวงพะยอม แก้วพิลา และคณะ (2549) ที่ศึกษาถึงข้อสอบที่ใช้สอบรวบยอดพบว่า ข้อสอบส่วนใหญ่อยู่ในระดับความยาก-ง่ายปานกลาง ถึงง่าย ซึ่งสอดคล้องกับวัตถุประสงค์ของการสอบรวบยอดคือ มีความรู้ในระดับที่จำเป็น เพื่อให้สำเร็จการศึกษา มีความเที่ยงของข้อสอบ อยู่ในช่วง 0.5-0.6 ซึ่งมีค่าความเชื่อมั่นค่อนข้างต่ำ ซึ่งในงานวิจัยดังกล่าวได้มีการลดค่าความคลาดเคลื่อนของการตัดสินผลโดยเพิ่ม ความกว้างของค่าความเชื่อมั่น (confident interval) และในการศึกษานี้ มีการหาความสัมพันธ์ระหว่างค่าความยาก-ง่ายของข้อสอบก่อนสอบกับค่าความยาก-ง่ายของข้อสอบหลังการสอบ พบว่ามีความสัมพันธ์กัน จำนวน 353 ข้อ คิดเป็นร้อยละ 88.25 เป็นที่ยอมรับได้ มีความเหมาะสมสำหรับเนื้อหาที่สำคัญต้องรู้ ควรรู้และน่ารู้ ตามลำดับ มีความสอดคล้องกับแนวคิดและทฤษฎีที่กำหนดไว้ในงานวิจัย และจากการศึกษาของธนาภรณ์ แดงใจและคณะ (2561) ทำการศึกษาเรื่อง ความสอดคล้องของค่า AI กับผลการสอบ MCQ ของนักศึกษาแพทย์ชั้นปีที่ 5 รายวิชาเวชศาสตร์ปฐมภูมิ 2 คณะแพทยศาสตร์โรงพยาบาลรามาธิบดี มหาวิทยาลัยมหิดล พบว่า ข้อสอบชุดที่ 2 เป็นชุดข้อสอบที่มีค่า AI สัมพันธ์ทางบวกกับค่า p เป็นชุดข้อสอบที่มีค่าอำนาจจำแนกของข้อสอบในระดับดีมาก ส่วนข้อสอบชุดที่ 1, 3 และ 4 มีค่า AI ที่ไม่สัมพันธ์กับค่า p ที่แสดงให้เห็นว่าการวิเคราะห์ค่า AI ของคณะกรรมการวิเคราะห์ข้อสอบยังไม่สอดคล้องกับความยาก-ง่ายของข้อสอบและผลการสอบ

จากการศึกษานี้ พบว่าข้อสอบชุดที่ 4 เป็นข้อสอบชุดที่ได้รับการเลือกนำไปใช้สอบ กับนิสิตแพทย์เพียง 13 คน ค่าความเชื่อมั่นของชุดข้อสอบชุดดังกล่าว อยู่ที่ 0.821 ซึ่งเป็นที่ยอมรับได้ แต่พบว่าค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สันไม่มีความสัมพันธ์กันทางสถิติ ทำให้ไม่สามารถนำผลวิเคราะห์มาตัดสินได้ สอดคล้องกับการกล่าวถึงเทคนิคร้อยละ 27 ซึ่งเหมาะกับการใช้คำนวณค่าความยาก-ง่าย และอำนาจจำแนกด้วยเทคนิคดังกล่าว กับข้อสอบแบบตัวเลือกที่ควรมีผู้สอบตั้งแต่ 30 คนขึ้นไป (จักรพงษ์ แผ่นทอง, 2564) จึงควรมีการนำข้อสอบชุดดังกล่าวไปใช้สอบกับนิสิตแพทย์ในจำนวนที่มากขึ้น เพื่อทดสอบค่ามาตรฐานต่าง ๆ ให้แสดงถึงคุณภาพของข้อสอบชุดนี้ต่อไป

จากการศึกษานี้ พบจุดแข็งของข้อสอบที่คณะกรรมการรายวิชา มีความตั้งใจในการจัดทำชุดข้อสอบขึ้นมาเพื่อใช้ในการจัดสอบให้เป็นมาตรฐานร่วมกันกับเครือข่าย ซึ่งได้เกิดจากความร่วมมือร่วมใจจากอาจารย์จากทุกศูนย์แพทย์ ใช้เวลาในการจัดทำข้อสอบทั้ง 400 ข้อนี้ ผ่านกระบวนการในการออกข้อสอบ และนำข้อสอบมาประชุมกลุ่มกรองข้อสอบร่วมกัน โดยยึดถึงวัตถุประสงค์ที่รายวิชาได้กำหนดไว้ และประโยชน์ของผู้สอบเป็นสำคัญ โดยการกำหนดค่า AI ก่อนการสอบว่าผู้สอบกลุ่มก่อนจะสามารถทำข้อสอบข้อนั้น ๆ ได้ และออกข้อสอบตามเกณฑ์ของแพทยสภา ซึ่งเป็นเกณฑ์ที่นิสิตแพทย์จำเป็นต้องรู้ และสามารถทำได้

จากการศึกษานี้ พบว่ายังมีจุดอ่อนด้านการเลือกใช้ชุดข้อสอบ ความถี่ในการเลือกใช้ชุดข้อสอบในชุดที่ 4 ที่มีการเลือกใช้เพียง 1 ครั้ง และเลือกใช้กับกลุ่มผู้สอบ เพียง 13 คน ทำให้อาจได้ผลการวิเคราะห์ที่ไม่เป็นไปตามมาตรฐาน อีกทั้งยังพบว่าการเลือกใช้ข้อสอบชุดที่ 1 มีการเลือกใช้ถึง 2 ครั้ง ซึ่งอาจทำให้เกิดความแตกต่างกันของชุดข้อสอบทั้ง 2 ชุดนี้

จากการศึกษานี้ ควรมีการพิจารณาปรับปรุงข้อสอบทั้ง 4 ชุด และข้อสอบใหม่ เพื่อให้มีความยากง่ายได้มาตรฐานยิ่งขึ้นและสามารถประเมินผลการเรียนได้อย่างถูกต้องแม่นยำมากขึ้น รวมถึงการพิจารณาค่า AI ของข้อสอบในแต่ละข้อให้มีความเหมาะสมยิ่งขึ้น โดยควรดำเนินการดังนี้

1. ส่งข้อสอบและผลการวิเคราะห์ข้อสอบให้ผู้ออกข้อสอบทราบข้อบกพร่องเพื่อปรับปรุง
2. นำข้อสอบที่ได้ปรับปรุงแล้วมาเข้าที่ประชุมคณะกรรมการรายวิชาวิเคราะห์ข้อสอบใหม่อีกครั้งและพิจารณา
ค่า AI ที่เหมาะสม
3. นำข้อสอบที่ได้มาตรฐานแล้วเก็บเข้าคลังข้อสอบเพื่อใช้ในปีการศึกษาถัดไป
4. ข้อสอบที่เหลือจากการจัดทำชุดข้อสอบ ควรให้คณะกรรมการรายวิชาวิเคราะห์ข้อสอบนำมาพิจารณาใช้ในปี
การศึกษาต่อไป เพื่อวิเคราะห์และปรับปรุงหาข้อสอบที่ได้มาตรฐานมาเก็บไว้ในคลังข้อสอบเพิ่มเติม
5. วิเคราะห์ผลการสอบโดยการหาค่า P เพื่อช่วยในการพิจารณาหาความเหมาะสมของค่า AI ต่อไป
6. จัดอบรมการออกข้อสอบและการวิเคราะห์ข้อสอบที่ได้มาตรฐานให้กับอาจารย์ใหม่ และอาจารย์เก่าที่ยังไม่
เคยได้รับการอบรมดังกล่าว เพื่อให้อาจารย์ตระหนักถึงความสำคัญของข้อสอบในการประเมินผลการวัดมาตรฐานของ
คณะแพทยศาสตร์ มหาวิทยาลัยนเรศวร ต่อไปในอนาคต

ข้อเสนอแนะ

1. เสนอให้คณะกรรมการรายวิชาพิจารณาปรับปรุงข้อสอบที่มีค่าอำนาจการจำแนกข้อสอบต่ำ เพื่อเพิ่ม
ประสิทธิภาพในการประเมินผลการเรียนรู้
2. เสนอให้คณะกรรมการรายวิชาพิจารณาปรับปรุงแก้ไขข้อสอบรวมทั้งการพิจารณาถึงค่า AI ของข้อสอบ
ทั้ง 4 ชุด เพื่อให้มีความง่ายที่ได้มาตรฐาน และสามารถประเมินผลการเรียนได้อย่างถูกต้องแม่นยำมากยิ่งขึ้น
3. การเลือกชุดข้อสอบที่นำไปใช้กับนิสิตแพทย์ ควรให้มีจำนวนผู้เข้าสอบมากขึ้นในข้อสอบชุดที่ 4 เพื่อการ
ทดสอบค่ามาตรฐานต่าง ๆ ที่แสดงถึงคุณภาพของชุดข้อสอบที่มีมาตรฐานต่อไป

เอกสารอ้างอิง

- มหาวิทยาลัยรามคำแหง. **การวิเคราะห์ข้อสอบ**. สืบค้น 21 พฤษภาคม 2567, จาก
<http://old-book.ru.ac.th/e-book/m/MR311/mr311-10.pdf>.
- จรวงศ์ ประดับศรี. (2550). **คุณภาพของแบบทดสอบที่มีรูปแบบตัวเลือกและวิธีวิเคราะห์ต่างกัน**. (วิทยานิพนธ์
ปริญญาวิทยาศาสตรมหาบัณฑิต). สงขลา: มหาวิทยาลัยราชภัฏสงขลา.
- จักรพงษ์ แผ่นทอง. (2564). **การหาอำนาจจำแนกแบบทดสอบด้วยเทคนิค ร้อยละ 27 หรือวิธีของจุง เทห์ ฟาน
(Chung-Teh Fan)**. สืบค้น 3 กุมภาพันธ์ 2568, จาก <https://krujakrapong.com>.
- เชิดศักดิ์ ไอรณรัตน์. (2552). **การวิเคราะห์ข้อสอบปรนัย ภาควิชาศัลยศาสตร์ คณะแพทยศาสตร์ ศิริราชพยาบาล
มหาวิทยาลัยมหิดล**, *เวชบันทึกศิริราช*, 2(1), 31-37.
- ธนภรณ์ แดงใจ, ปณิธิ พูนเพชรรัตน์, จิรวดี ชุมพล และ ทิชากร อชาติตวรากุล. (2561). **ความสอดคล้องของค่า
Acceptability Index กับผลการสอบ MCQ ของนักศึกษาแพทย์ชั้นปีที่ 5 วิชาเวชศาสตร์ปฐมภูมิ 2
คณะแพทยศาสตร์ โรงพยาบาลรามคำแหง มหาวิทยาลัยมหิดล**, *รามคำแหงวารสาร*, 41(1), 1-8.
- ล้วน สายยศ และ อังคนา สายยศ. (2543). **เทคนิคการวัดผลการเรียนรู้** (พิมพ์ครั้งที่ 2). กรุงเทพฯ: สุวีริยาสาส์น.
- Downing, S. M. (2003). Validity: on meaningful interpretation of assessment data. *Med Educ*, 37(9),
830-837. from <https://doi.org/10.1046/j.1365-2923.2003.01594.x>.

Downing, S. M. (2004). Reliability: on the reproducibility of assessment data. **Med Educ**, **38**(9), 1006–1012. from <https://doi.org/10.1111/j.1365-2929.2004.01932.x>.

Kaewpila, P., Chau-In, W., Limratana, N., & Panthongviriyakul, C. (2006). Post-test analysis of the comprehensive examination of Thailand Medical School Consortium for the 2005 KKU sixth-year students. **Srinagarind Medical Journal**, **21**(3), 200–206.