

การประเมินผลการอบรม “การพัฒนาข้อสอบปรนัยเพื่อประเมินความรู้ทางการแพทย์”
ของศูนย์ความเป็นเลิศด้านการศึกษาวิทยาศาสตร์สุขภาพ
โดยใช้รูปแบบการประเมินเคิร์กแพทริก

Evaluation of a Workshop “Multiple-Choice Question Item Development
for Assessing Medical Knowledge”, by Siriraj Health Science Education
Excellence Center Using Kirkpatrick Model

Received:	March	9, 2023
Revised:	May	14, 2023
Accepted:	June	12, 2023

ภัทรพร นาคนาเกร็ด (Pattaraporn Naknagraed)^{*}
กนกพล ศุภสิริมนตรี (Kanokphol Supasirimontri)^{**}
เชิดศักดิ์ ไอรณณรัตน์ (Cherdsak Iramaneerat)^{***}

บทคัดย่อ

“การพัฒนาข้อสอบปรนัยเพื่อประเมินความรู้ทางการแพทย์” จัดโดยศูนย์ความเป็นเลิศด้านการศึกษาวิทยาศาสตร์สุขภาพศิริราช ระหว่างวันที่ 21-22 มิถุนายน 2564 วัตถุประสงค์การวิจัย คือ ประเมินการอบรมการพัฒนาข้อสอบปรนัยเพื่อประเมินความรู้ทางการแพทย์ โดยใช้รูปแบบเคิร์กแพทริก

วิธีการศึกษา ขอบเขตการวิจัยอิงตามรูปแบบเคิร์กแพทริก 4 ระดับ แบ่งเป็น 2 ระยะ ได้แก่ ระหว่างอบรม และ 3 เดือนหลังสิ้นสุดอบรม ประชากร คือ ผู้เข้าอบรม จำนวน 36 คน และผู้บังคับบัญชา เครื่องมือที่ใช้คือ แบบประเมิน แบบสอบถาม และแบบทดสอบ ประกอบด้วย 1) ระดับ 1 ปฏิบัติการ แบบประเมินความพึงพอใจของผู้อบรม 2) ระดับ 2 การเรียนรู้ แบบทดสอบก่อนและหลังอบรมเพื่อประเมินความรู้ 3) ระดับ 3 พฤติกรรมแบบสอบถามประเมินทักษะของการพัฒนาข้อสอบปรนัย 4) ระดับ 4 ผลลัพธ์ แบบสอบถามเพื่อประเมินผลหลังสิ้นสุดการอบรม สถิติที่ใช้ คือ ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน และค่า t-test

ผลการศึกษา ระดับ 1 ปฏิบัติการ ค่าเฉลี่ยความพึงพอใจต่อหัวข้ออบรมและบริหารจัดการ เท่ากับ 4.64, 4.56 ตามลำดับ ระดับ 2 การเรียนรู้ คะแนนความรู้หลังจากอบรมสูงกว่าคะแนนก่อนอบรมอย่างมีนัยสำคัญทางสถิติ ($t = 8.50, p < 0.05$) ระดับ 3 พฤติกรรม ผู้อบรมนำความรู้ที่ได้รับจากการอบรมไปปฏิบัติบ่อยครั้ง ($\bar{X} = 3.91$) และผู้บังคับบัญชาเห็นว่าผู้อบรมนำความรู้ไปปฏิบัติเป็นประจำ ($\bar{X} = 4.51$) ระดับ 4 ผลลัพธ์โดยรวม ผู้อบรมและผู้บังคับบัญชาเห็นด้วยว่ามีประโยชน์ต่อสถาบัน ครู และผู้เรียนมาก ($\bar{X} = 4.49, 4.46$ ตามลำดับ) จากเกณฑ์ตัดสินพบว่า ผลการประเมินผ่านทุกด้าน

^{*} นักวิชาการศึกษา คณะแพทยศาสตร์ศิริราชพยาบาล มหาวิทยาลัยมหิดล / pattaraporn.nak@mahidol.edu
Educator, Faculty of Medicine Siriraj Hospital, Mahidol University / pattaraporn.nak@mahidol.edu

^{**} นพ. คณะแพทยศาสตร์ศิริราชพยาบาล มหาวิทยาลัยมหิดล / kanokphol.sup@mahidol.edu
Dr., Faculty of Medicine Siriraj Hospital, Mahidol University / kanokphol.sup@mahidol.edu

^{***} รองศาสตราจารย์ ดร. นพ. คณะแพทยศาสตร์ศิริราชพยาบาล มหาวิทยาลัยมหิดล / cherdsak.ira@mahidol.ac.th
Associate Professor, Dr., Faculty of Medicine Siriraj Hospital, Mahidol University / cherdsak.ira@mahidol.ac.th

สรุปผลการศึกษา โดยการประเมินแต่ละด้านสะท้อนถึงประสิทธิภาพของการอบรม ในด้านปฏิริยา
หากการอบรมมีการเตรียมหัวข้ออบรมที่เหมาะสมและมีการบริหารจัดการที่ดี ย่อมจะส่งผลต่อพฤติกรรม
การเรียนรู้และปฏิบัติของผู้อบรม ทำให้ส่งผลดีต่อสถาบัน ครู และผู้เรียนด้วย

คำสำคัญ : การประเมินผลหลักสูตรอบรม, ข้อสอบปรนัย, รูปแบบการประเมินเคิร์กแพทริค

Abstract

A “Developing Multiple-Choice Questions (MCQs) for assessing medical knowledge” was organized by Siriraj Health science Education Excellence center (SHEE) during the 21st-22nd June 2021. Objective of this research was to evaluate this workshop using Kirkpatrick Model.

Methods: Scope of this research based on four levels of the Kirkpatrick Model, divided into 2 phases: during workshop and 3 months after the workshop. The populations were 36 workshop participants and their supervisors. The instrumentation consisted of 1) level 1 (reaction), participants’ satisfaction questionnaire 2) level 2 (learning), pre-test and post-test for assessing knowledge 3) level 3 (behavior), questionnaire to assess the skills of MCQs development 4) level 4 (results), questionnaire to assess the outcomes after the end of workshop. Statistics used were mean, standard deviation, t-test.

Results: Level 1 (reaction): the means of satisfaction with the training topic and management were 4.64, and 4.56, respectively. Level 2 (learning): the scores after training were statistically significantly higher than the scores before training ($t = 8.50$, $p < 0.05$). Level 3 (behavior): the major opinion from trainees was the knowledge gained from the training was often applied at the practical ($\bar{X} = 3.91$) and the supervisors agreed that the participants frequently used the knowledge in their practices ($\bar{X} = 4.51$) Level 4 (results): Overall, the trainees and supervisors agreed that there was a high of benefit to the institution, (teachers and students $\bar{X} = 4.49$, 4.46 , respectively). According to the judging criteria, it was found that the assessment results were at a pass level in all aspects.

Conclusion: The assessment in each area will reflect the effectiveness of the training. On the reaction side, if the training topic is prepared and appropriate for good management, it will affect the practical learning and behavior of participants and proceed to good results for the institution, teachers, and students as well.

Keywords: Training course evaluation, Multiple-choice question, Kirkpatrick evaluation model

บทนำ

การสร้างและพัฒนาข้อสอบปรนัยในโรงเรียนแพทย์ เพื่อประเมินความรู้ความเข้าใจของนักศึกษาแพทย์นั้น “อาจารย์แพทย์” ที่ทำหน้าที่สอนมีความสำคัญอย่างยิ่งในกระบวนการดังกล่าว แต่ยังคงพบว่าในโรงเรียนแพทย์มีอาจารย์แพทย์ผู้สอนจำนวนไม่น้อยมีปัญหาเกี่ยวกับการสร้างและพัฒนาข้อสอบปรนัย โดยข้อสอบปรนัยที่สร้างและพัฒนาขึ้นอย่างดีและถูกวิธีนั้น สามารถวัดความรู้ได้ทั้งระดับการจดจำ การทำความเข้าใจ และการประยุกต์ความรู้ไปใช้ในการดูแลคนไข้ แต่หากข้อสอบปรนัยที่ถูกสร้างและพัฒนาขึ้นอย่างไม่ถูกหลักการจะส่งผลเสียหลายอย่าง เช่น ทำให้ข้อสอบยากขึ้นโดยไม่จำเป็น ทำให้ผู้สอบเกิดความสับสน ทำให้ผู้สอบบางกลุ่มเสียเปรียบผู้สอบคนอื่น ทำให้การตัดสินผลสอบผิดพลาด เป็นต้น ดังนั้นจึงมีความจำเป็นอย่างยิ่งที่อาจารย์แพทย์ต้องมีความรู้พื้นฐานเกี่ยวกับขั้นตอนและแนวทางการสร้างข้อสอบปรนัย ข้อควรระวังในการสร้างข้อสอบปรนัย การวิเคราะห์ข้อสอบ (item analysis) และการตั้งเกณฑ์ผ่านการสอบ (passing standard) เพื่อให้ได้ข้อสอบปรนัยที่ดีมีคุณภาพ ด้วยเหตุผลนี้ ศูนย์ความเป็นเลิศด้านการศึกษาวิทยาศาสตร์สุขภาพ (Siriraj Health Science Education Excellence center: SHEE) คณะแพทยศาสตร์ศิริราชพยาบาล จึงมีการจัดหลักสูตรอบรม “การพัฒนาข้อสอบปรนัยเพื่อประเมินความรู้ทางการแพทย์” อย่างต่อเนื่องทุกปี ตั้งแต่ปี พ.ศ. 2554 ถึงปัจจุบัน โดยมีวัตถุประสงค์ คือ ให้ผู้อบรมมีความรู้และเข้าใจหลักการพื้นฐานของข้อสอบปรนัย การวางแผนและจัดทำตารางกำหนดเนื้อหาข้อสอบ การสร้างโจทย์และตัวเลือกข้อสอบปรนัย การพิจารณาข้อสอบ การผลิตข้อสอบและการจัดสอบ การตรวจและวิเคราะห์ข้อสอบปรนัย การจัดทำคลังข้อสอบ การตั้งเกณฑ์ผ่านการสอบ และการตัดเกรด ซึ่งเป็นเนื้อหาที่จำเป็นสำหรับอาจารย์อย่างมาก เพราะข้อสอบปรนัย (Multiple-Choice Question: MCQ) เป็นรูปแบบการประเมินผลที่นิยมใช้ในวงการแพทยศาสตรศึกษา เนื่องด้วยคุณสมบัติที่ดีหลายประการ ได้แก่ ประสิทธิภาพของการประเมินความรู้ในปริมาณมากโดยใช้เวลาน้อย โดยการประเมินไม่มีผลจากความรู้สึกส่วนตัวของผู้ประเมินหรือผู้ให้คะแนน จึงมีความเที่ยงตรงสูง รวมถึงมีผลการวิจัยจำนวนมากสนับสนุนความถูกต้องของผลประเมินด้วยข้อสอบปรนัย ดังนั้น หากในโรงเรียนแพทย์หรือวิทยาศาสตร์สุขภาพ มีการพัฒนาข้อสอบปรนัยอย่างดีก็จะสามารถวัดความรู้ ระดับการทำความเข้าใจ และการประยุกต์ใช้ความรู้ในการดูแลคนไข้ (Iramaneerat, 2012: 29) ซึ่งหลักสูตรอบรมดังกล่าว ยังไม่เคยมีการประเมินประสิทธิผล หรือมีกระบวนการติดตามผลลัพธ์ของการอบรมมาก่อน

การประเมินรูปแบบเคิร์กแพทริก นิยมใช้ประเมินโครงการหรือหลักสูตรอบรม โดย โดแนล แอล เคิร์กแพทริก (Kirkpatrick, 1994) ได้เสนอแนะการประเมินผล 4 ระดับ คือ 1) ปฏิกริยา (reaction) เป็นการประเมินการตอบสนองของผู้อบรมว่า รู้สึกหรือความคาดหวังกับการอบรมอย่างไร โดยที่เครื่องมือรวบรวมข้อมูล อาจเป็นการสัมภาษณ์ แบบสอบถาม ควรมีคำถามปลายเปิดเพื่อให้แสดงความคิดเห็น 2) การเรียนรู้ (learning) เป็นการประเมินความรู้ที่เกิดขึ้นกับผู้เข้าอบรม 3) พฤติกรรม (behavior) เป็นการประเมินพฤติกรรมเมื่อกลับไปปฏิบัติงานว่า เปลี่ยนแปลงที่พึงประสงค์หรือไม่ และประยุกต์ใช้ความรู้หลังจากอบรมมากน้อยเพียงไร เป็นต้น อาจใช้แบบสัมภาษณ์กับผู้บังคับบัญชา หรือแบบสังเกต 4) ผลลัพธ์ (results) ที่เกิดกับองค์กร เป็นการประเมินผลลัพธ์ที่เกิดผลกระทบต่อองค์กรอันเนื่องมาจากการเปลี่ยนแปลงพฤติกรรมของผู้อบรม อาจใช้การตรวจสอบแบบสอบถามผู้บังคับบัญชา (Panakul, 2002)

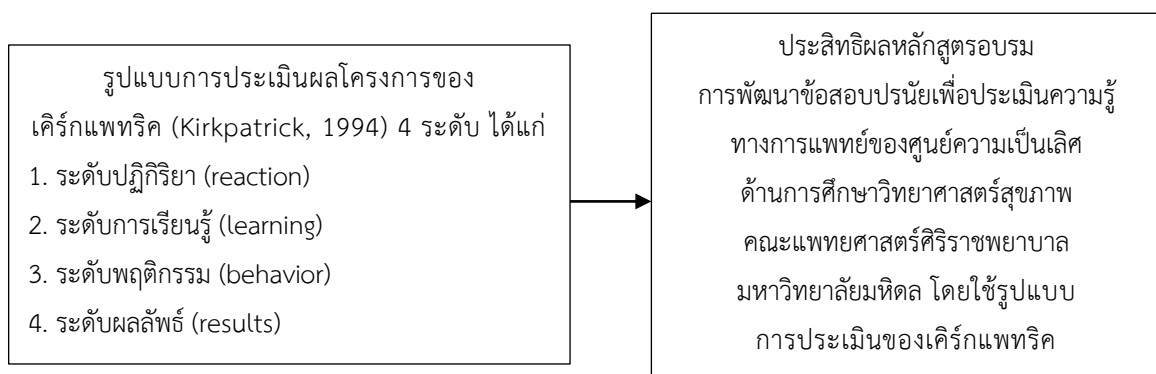
ด้วยความสำคัญของการวัดประเมินผล การพัฒนาข้อสอบปรนัยให้มีประสิทธิภาพ ผู้วิจัยจึงมีความสนใจที่จะประยุกต์แนวทฤษฎีของโดนัล แอล เคิร์กแพทริก (Donald L. Kirkpatrick) ในด้านปฏิกิริยาสะท้อนกลับ ด้านการเรียนรู้ ด้านพฤติกรรม ด้านผลลัพธ์ที่เกิดขึ้นต่อองค์กร มาใช้ในการประเมินผลหลักสูตรอบรมการพัฒนาข้อสอบปรนัยเพื่อประเมินความรู้ทางการแพทย์ของศูนย์ความเป็นเลิศด้านการศึกษาวิทยาศาสตร์สุขภาพ เพื่อประโยชน์ต่อการพัฒนาหลักสูตรอบรม และนำข้อมูลไปใช้ในระดับนโยบายต่อไป

วัตถุประสงค์ของการวิจัย

เพื่อประเมินผลหลักสูตรอบรมการพัฒนาข้อสอบปรนัยเพื่อประเมินความรู้ทางการแพทย์ของศูนย์ความเป็นเลิศด้านการศึกษาวิทยาศาสตร์สุขภาพ โดยใช้รูปแบบประเมินเคิร์กแพทริก

การศึกษาเอกสารที่เกี่ยวข้องและกรอบแนวคิดการวิจัย

การวิจัยนี้ใช้แนวทางการประเมินตามรูปแบบเคิร์กแพทริก 4 ระดับ (Kirkpatrick, 1994) และได้ทำการแบ่งระยะในการศึกษาออกเป็น 2 ระยะ คือ ระยะที่ 1 การประเมินระหว่างดำเนินโครงการ เพื่อประเมินปฏิกิริยาตอบสนองของผู้เข้ารับการอบรม และประเมินเปรียบเทียบผลการเรียนรู้ของผู้เข้ารับการอบรมระหว่างก่อนและหลังการฝึกอบรม และระยะที่ 2 การประเมินหลังสิ้นสุดโครงการ เพื่อประเมินพฤติกรรมที่เปลี่ยนไป และผลลัพธ์จากการอบรมในระดับสถาบัน อาจารย์หรือผู้สอน ผู้เรียน ดังแสดงตามภาพที่ 1



ภาพที่ 1 กรอบแนวคิดการวิจัย

นิยามศัพท์ที่ใช้ในการวิจัย

ประสิทธิผลหลักสูตรอบรม หมายถึง ผลลัพธ์ และผลกระทบที่ได้จากการดำเนินโครงการอบรม

หลักสูตรอบรมการพัฒนาข้อสอบปรนัยเพื่อประเมินความรู้ทางการแพทย์ของศูนย์ความเป็นเลิศด้านการศึกษาวิทยาศาสตร์สุขภาพ หมายถึง โครงการอบรมอาจารย์ของคณะแพทยศาสตร์ศิริราชพยาบาล โดยมีเนื้อหาเกี่ยวกับหลักการพื้นฐานของข้อสอบปรนัย การวางแผนและจัดทำตารางกำหนดเนื้อหาข้อสอบ การสร้างโจทย์และตัวเลือกข้อสอบปรนัย ข้อผิดพลาดที่พบบ่อยในการสร้างข้อสอบปรนัย การพิจารณาข้อสอบ การผลิตข้อสอบและการจัดสอบ การตรวจและวิเคราะห์ข้อสอบปรนัย การจัดทำคลังข้อสอบ การตั้งเกณฑ์ผ่านการสอบ และการตัดเกรด

รูปแบบการประเมินของเคิร์กแพทริก หมายถึง รูปแบบการประเมินที่ใช้ในการประเมินประสิทธิผลหลักสูตรอบรม การพัฒนาข้อสอบปรนัยเพื่อประเมินความรู้ทางการแพทย์ของศูนย์ความเป็นเลิศด้านการศึกษาวิทยาศาสตร์สุขภาพ คณะแพทยศาสตร์ศิริราชพยาบาล มหาวิทยาลัยมหิดล ประกอบด้วย 4 ระดับ คือ

การประเมินระดับปฏิกิริยา หมายถึง การประเมินความเหมาะสมของหลักสูตรอบรมว่าผู้ที่เข้ามารับการอบรมพอใจหรือไม่ต่อสิ่งที่ได้รับจากการอบรม ได้แก่ 1) หัวข้อการจัดอบรม ในด้านความรู้ เนื้อหาสาระประโยชน์ การนำไปประยุกต์ใช้ให้เกิดประโยชน์ เทคนิคการถ่ายทอดของวิทยากร การตอบข้อซักถามของวิทยากร ความเหมาะสม เวลา เอกสารประกอบ และ 2) การบริหารจัดการอบรม ในด้านสถานที่ ระยะเวลา อาหาร การประสานงาน ภาพรวมของการจัดอบรม โดยใช้แบบประเมินความพึงพอใจ

การประเมินระดับการเรียนรู้ หมายถึง การประเมินผลการเรียนรู้ของผู้เข้าอบรม โดยใช้แบบทดสอบวัดผลการเรียนรู้

การประเมินระดับพฤติกรรมที่เปลี่ยนไป หมายถึง การประเมินว่าผู้เข้าอบรมได้ปรับเปลี่ยนพฤติกรรมในการพัฒนาข้อสอบ การจัดสอบหรือไม่ หลังจากสิ้นสุดการอบรม โดยใช้แบบสอบถามเพื่อประเมินทักษะการพัฒนาข้อสอบปรนัย

การประเมินระดับผลลัพธ์ที่เกิดขึ้น หมายถึง การประเมินผลจากการอบรมว่าได้เกิดผลดีต่อสถาบัน อาจารย์ และผู้เรียน หรือเกิดผลกระทบในทางที่ดีในลักษณะใดบ้าง มีคุณภาพดีขึ้นหรือไม่ ซึ่งใช้แบบสอบถามเพื่อประเมินผลลัพธ์ที่เกิดขึ้นหลังสิ้นสุดการอบรม

วิธีดำเนินการวิจัย

เป็นการวิจัยประเมินผลโดยเก็บข้อมูลเชิงปริมาณ ประชากร คือ ผู้อบรมหลักสูตรการพัฒนาข้อสอบปรนัยเพื่อประเมินความรู้ทางการแพทย์ของศูนย์ความเป็นเลิศด้านการศึกษาวิทยาศาสตร์สุขภาพ ปี พ.ศ. 2564 จำนวน 36 คน และผู้บังคับบัญชาของผู้อบรม

การกำหนดขนาดตัวอย่างขั้นต่ำ โดยใช้โปรแกรม G*power กำหนดการวิเคราะห์ข้อมูลที่ใช้สถิติ T-tests กำหนดค่า effect size เท่ากับ 0.80 ค่าความคลาดเคลื่อน (alpha) เท่ากับ 0.05 ค่า power เท่ากับ 0.80 ได้จำนวนกลุ่มตัวอย่าง 27 คน อย่างไรก็ตาม พบว่า ขนาดของกลุ่มตัวอย่างที่เหมาะสมและที่ทำให้เกิดการผิดพลาดในการศึกษาประมาณร้อยละ 50 ควรมีขนาดกลุ่มตัวอย่าง 40 คน โดยขนาดที่มากขึ้นจะทำให้มีความถูกต้องมากยิ่งขึ้น (Cohen, 1977) ดังนั้น จึงเก็บข้อมูลกับประชากรทั้งหมด โดยกำหนดเกณฑ์คัดออก คือ ถ้าผู้ให้ข้อมูลไม่สะดวกให้ข้อมูล และเกณฑ์การถอนหรือยุติการเข้าร่วมวิจัย คือ ความไม่สมบูรณ์ของการตอบ หาก Missing Data มากกว่า 5% ของแบบสอบถาม จะถอนข้อมูลจากวิเคราะห์

สถิติที่ใช้ในการศึกษาวิเคราะห์ข้อมูล

สถิติพื้นฐาน ได้แก่ ค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน ใช้สำหรับวิเคราะห์ระดับความพึงพอใจ ทักษะการพัฒนาข้อสอบปรนัย และผลลัพธ์ที่เกิดขึ้นหลังสิ้นสุดการอบรม

สถิติเชิงเปรียบเทียบ เพื่อทดสอบสมมติฐาน โดยวิเคราะห์เปรียบเทียบคะแนนทดสอบก่อนอบรมและหลังอบรม ใช้สถิติ t-test แบบกลุ่มตัวอย่างไม่เป็นอิสระ (t-test dependent group)

เครื่องมือที่ใช้ในการวิจัย

ระยะที่ 1 การประเมินในระหว่างดำเนินโครงการ

1. การประเมินปฏิกิริยาของผู้เข้ารับการอบรม

1.1 เครื่องมือที่ใช้ในการวิจัย เครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูลการประเมินปฏิกิริยาตอบสนองของผู้เข้ารับการอบรม คือ แบบประเมินความพึงพอใจของผู้เข้าอบรมต่อโครงการอบรม “การพัฒนาข้อสอบปรนัยเพื่อประเมินความรู้ทางการแพทย์” ซึ่งเป็นข้อมูลทุติยภูมิ (secondary data) มาดำเนินการวิเคราะห์โดยแบบประเมินความพึงพอใจจะมีรายละเอียด ดังนี้

ตอนที่ 1 เป็นคำถามความพึงพอใจต่อหัวข้อการจัดอบรม ได้แก่ ด้านความรู้ เนื้อหาสาระประโยชน์ การนำไปประยุกต์ใช้ให้เกิดประโยชน์ เทคนิคการถ่ายทอดของวิทยากร การตอบข้อซักถามของวิทยากร ความเหมาะสม เวลา เอกสารประกอบ

ตอนที่ 2 เป็นคำถามความพึงพอใจต่อการบริหารจัดการอบรม ได้แก่ สถานที่ ระยะเวลา อาหาร การประสานงาน และภาพรวมของการจัดอบรม

ตอนที่ 3 เป็นคำถามปลายเปิดเกี่ยวกับสิ่งที่ท่านประทับใจในการฝึกอบรมและข้อเสนอแนะอื่น ๆ

โดยแบบประเมินตอนที่ 1 และตอนที่ 2 เป็นข้อมูลการประเมินเกี่ยวกับความพึงพอใจ โดยใช้มาตรวัด 5 ระดับของลิเคิร์ต (Likert) ให้ผู้ตอบแบบประเมินเลือกแสดงความคิดเห็นจากข้อความที่ผู้วิจัยได้กำหนดไว้ โดยกำหนดคะแนนความต้องการไว้ดังนี้

มีความรู้สึกเห็นด้วยในระดับมากที่สุด	ให้คะแนนเท่ากับ	5
มีความรู้สึกเห็นด้วยในระดับมาก	ให้คะแนนเท่ากับ	4
มีความรู้สึกเห็นด้วยในระดับปานกลาง	ให้คะแนนเท่ากับ	3
มีความรู้สึกเห็นด้วยในระดับน้อย	ให้คะแนนเท่ากับ	2
มีความรู้สึกเห็นด้วยในระดับน้อยที่สุด	ให้คะแนนเท่ากับ	1

1.2 การวิเคราะห์ข้อมูล การวิเคราะห์ข้อมูลปฏิกิริยาการตอบสนองของผู้เข้ารับการอบรมด้วยค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน โดยเกณฑ์มีการแปลผล ดังนี้ (Silpcharu, 2006: 77)

คะแนนเฉลี่ย 4.50 ถึง 5.00	หมายถึง	มีความรู้สึกพึงพอใจในระดับมากที่สุด
คะแนนเฉลี่ย 3.50 ถึง 4.49	หมายถึง	มีความรู้สึกพึงพอใจในระดับมาก
คะแนนเฉลี่ย 2.50 ถึง 3.49	หมายถึง	มีความรู้สึกพึงพอใจในระดับปานกลาง
คะแนนเฉลี่ย 1.50 ถึง 2.49	หมายถึง	มีความรู้สึกพึงพอใจในระดับน้อย
คะแนนเฉลี่ย 1.00 ถึง 1.49	หมายถึง	มีความรู้สึกพึงพอใจในระดับน้อยที่สุด

2. การประเมินการเรียนรู้ของผู้เข้าอบรม

2.1 เครื่องมือที่ใช้ในการวิจัย เครื่องมือที่ใช้ในการประเมินการเรียนรู้ของผู้เข้าอบรม คือ แบบทดสอบ เป็นแบบทดสอบวัดความรู้ความเข้าใจของผู้เข้าอบรมระหว่างก่อนและหลังการอบรมเกี่ยวกับหลักการพื้นฐานของข้อสอบปรนัย การวางแผนและจัดทำตารางกำหนดเนื้อหาข้อสอบ การสร้างโจทย์และตัวเลือก

ข้อสอบปรนัย ข้อผิดพลาดที่พบบ่อยในการสร้างข้อสอบปรนัย การพิจารณาข้อสอบ การผลิตข้อสอบและการจัดสอบ การตรวจและวิเคราะห์ข้อสอบปรนัย การจัดทำคลังข้อสอบ การตั้งเกณฑ์ผ่านการสอบ การตัดเกรด ลักษณะแบบทดสอบชนิดเลือกตอบ 5 ตัวเลือก จำนวน 30 ข้อ ซึ่งเป็นข้อมูลทุติยภูมิ (secondary data) มาดำเนินการวิเคราะห์

2.2 การวิเคราะห์ข้อมูล การวิเคราะห์ข้อมูลผลการเรียนรู้ของผู้เข้าอบรมระหว่างก่อนและหลังการอบรมใช้สถิติค่าที่แบบไม่อิสระจากกัน (t-test dependent)

ระยะที่ 2 การประเมินหลังจากโครงการสิ้นสุดไปแล้ว

1. การประเมินพฤติกรรมที่เปลี่ยนไปหลังการอบรม

1.1 เครื่องมือที่ใช้ในการวิจัย เครื่องมือที่ใช้ในการประเมินพฤติกรรมที่เปลี่ยนไปหลังการอบรม ได้แก่ แบบสอบถาม เพื่อประเมินทักษะการพัฒนาข้อสอบปรนัย โดยให้ผู้บังคับบัญชาประเมินพฤติกรรมที่เปลี่ยนไปของผู้เข้าอบรม และผู้เข้าอบรมประเมินตนเอง โดยทักษะการออกข้อสอบ จะเป็นมาตรวัด 5 ระดับ ดังนี้

5 แทน ปฏิบัติเป็นประจำ	= 5 คะแนน
4 แทน ปฏิบัติบ่อยครั้ง	= 4 คะแนน
3 แทน ปฏิบัติเป็นบางครั้ง	= 3 คะแนน
2 แทน ปฏิบัตินาน ๆ ครั้ง	= 2 คะแนน
1 แทน ไม่เคยปฏิบัติ	= 1 คะแนน

1.2 การวิเคราะห์ข้อมูล การวิเคราะห์ข้อมูลแบบสอบถาม เพื่อประเมินทักษะการพัฒนาข้อสอบปรนัย ด้วยการหาค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน โดยมีเกณฑ์การแปลผลดังนี้ (Silpcharu, 2006: 77)

คะแนนเฉลี่ย 4.51 – 5.00	หมายถึง	ปฏิบัติเป็นประจำ
คะแนนเฉลี่ย 3.51 – 4.50	หมายถึง	ปฏิบัติบ่อยครั้ง
คะแนนเฉลี่ย 2.51 – 3.50	หมายถึง	ปฏิบัติเป็นบางครั้ง
คะแนนเฉลี่ย 1.51 – 2.50	หมายถึง	ปฏิบัตินาน ๆ ครั้ง
คะแนนเฉลี่ย 1.00 – 1.50	หมายถึง	ไม่เคยปฏิบัติ

2. ประเมินผลลัพธ์ที่เกิดขึ้นหลังสิ้นสุดการอบรม

2.1 เครื่องมือที่ใช้ในการวิจัย เครื่องมือที่ใช้ในการประเมินผลลัพธ์ที่เกิดขึ้นกับสถาบัน อาจารย์ และผู้เรียน หลังสิ้นสุดการอบรม ได้แก่ แบบสอบถามเพื่อประเมินผลลัพธ์ที่เกิดขึ้นหลังสิ้นสุดการอบรม ซึ่งมีรายละเอียดดังนี้

ตอนที่ 1 เป็นรายละเอียดข้อมูลทั่วไปเกี่ยวกับสถานะ สังกัด

ตอนที่ 2 เป็นคำถามเกี่ยวกับผลลัพธ์ที่เกิดขึ้นกับสถาบัน อาจารย์ และผู้เรียน โดยใช้มาตรวัด 5 ระดับของลิเคิร์ต (Likert) ให้ผู้ตอบแบบสอบถามเลือกแสดงความคิดเห็นจากข้อความที่ผู้วิจัยได้กำหนดไว้ โดยมีระดับความคิดเห็นออกเป็น 5 ระดับ ดังนี้

มีความรู้สึกเห็นด้วยในระดับมากที่สุด	ให้คะแนนเท่ากับ	5
มีความรู้สึกเห็นด้วยในระดับมาก	ให้คะแนนเท่ากับ	4
มีความรู้สึกเห็นด้วยในระดับปานกลาง	ให้คะแนนเท่ากับ	3
มีความรู้สึกเห็นด้วยในระดับน้อย	ให้คะแนนเท่ากับ	2
มีความรู้สึกเห็นด้วยในระดับน้อยที่สุด	ให้คะแนนเท่ากับ	1

ตอนที่ 3 เป็นคำถามปลายเปิดเกี่ยวกับความคิดเห็นและข้อเสนอแนะ

2.2 การวิเคราะห์ข้อมูล การวิเคราะห์ข้อมูลผลลัพธ์ที่เกิดหลังการอบรม ด้วยค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน โดยเกณฑ์มีการแปลผลดังนี้ (Silpcharu, 2006: 77)

คะแนนเฉลี่ย 4.50 ถึง 5.00	หมายถึง	มีความรู้สึกเห็นด้วยในระดับมากที่สุด
คะแนนเฉลี่ย 3.50 ถึง 4.49	หมายถึง	มีความรู้สึกเห็นด้วยในระดับมาก
คะแนนเฉลี่ย 2.50 ถึง 3.49	หมายถึง	มีความรู้สึกเห็นด้วยในระดับปานกลาง
คะแนนเฉลี่ย 1.50 ถึง 2.49	หมายถึง	มีความรู้สึกเห็นด้วยในระดับน้อย
คะแนนเฉลี่ย 1.00 ถึง 1.49	หมายถึง	มีความรู้สึกเห็นด้วยในระดับน้อยที่สุด

การเก็บรวบรวมข้อมูล

1. ศึกษาทฤษฎี งานวิจัยในอดีตเกี่ยวกับการหาประสิทธิภาพโครงการ
2. สร้างเครื่องมือที่ใช้ในการวิจัย
3. ตรวจสอบความตรงตามเนื้อหากับผู้ทรงคุณวุฒิ จำนวน 4 คน โดยกำหนดที่ค่า IOC สูงกว่า 0.5 ถือว่าเป็นข้อคำถามที่ยอมรับให้นำไปใช้เก็บข้อมูลได้
4. นำเครื่องมือที่ผ่านการพัฒนาเรียบร้อยแล้วไปใช้จริงกับกลุ่มตัวอย่าง โดยมีเอกสารชี้แจงกับกลุ่มตัวอย่าง เพื่อสร้างความเข้าใจและชี้แจงให้กลุ่มตัวอย่างเกิดความเข้าใจจุดประสงค์ของการวิจัย และกระบวนการวิจัยว่าจะไม่มีผลกระทบต่อค่าบริการและมาตรฐานของการอบรมแต่ประการใด และสามารถเลือกเข้าร่วมหรือไม่เข้าก็ได้ โดยให้กลุ่มตัวอย่างระบุชื่อผู้บังคับบัญชาของตน หรือผู้บังคับบัญชาที่ดูแลการสอบโดยตรงเพื่อทางผู้วิจัยจะส่งแบบสอบถามไปยังผู้บังคับบัญชาเมื่อสิ้นสุดการอบรมแล้ว 3 เดือน
5. ดำเนินการทดสอบความรู้ก่อนการอบรมกับกลุ่มตัวอย่าง จากนั้นกลุ่มตัวอย่างจะเข้าสู่กระบวนการอบรมในหลักสูตร การพัฒนาข้อสอบปรนัยเพื่อประเมินความรู้ทางการแพทย์ ของคณะแพทยศาสตร์ศิริราชพยาบาล มหาวิทยาลัยมหิดล และเมื่อเสร็จสิ้นกระบวนการอบรม กลุ่มตัวอย่างจะทำการทดสอบความรู้หลังอบรมอีกครั้ง โดยเป็นแบบทดสอบชุดเดียวกันแต่สลับข้อ และสลับตัวเลือก
6. ดำเนินการแจกแบบสอบถามความพึงพอใจกับกลุ่มตัวอย่าง
7. ผู้วิจัยรวบรวมข้อมูล และวิเคราะห์ผลความพึงพอใจและผลการทดสอบความรู้ก่อนและหลังการอบรม

โดยสรุปผู้วิจัยมีการดำเนินการวิจัย โดยใช้รูปแบบการประเมินตามรูปแบบเคิร์กแพทริก 4 ระดับ (Kirkpatrick, 1994) มีการประเมินระหว่างอบรม และหลังอบรม โดยใช้เครื่องมือวิจัยที่วัดผลในแต่ละระดับ ดังแสดงตามตารางที่ 1

ตารางที่ 1 กรอบการดำเนินการวิจัย โดยใช้รูปแบบการประเมินของเคิร์กแพทริก

	วัตถุประสงค์	ประเด็นที่ศึกษา	ตัวบ่งชี้	แหล่งข้อมูล	เครื่องมือที่ใช้	การวิเคราะห์และเกณฑ์ตัดสิน
ระยะที่ 1 ประเมินระหว่างจัดอบรม	1. ด้านปฏิกิริยา (reaction)	ความพึงพอใจต่อหลักสูตรอบรมฯ	- ด้านหัวข้อการอบรม ประเด็นความรู้ เนื้อหา การประยุกต์ใช้ให้เกิดประโยชน์ เทคนิคการถ่ายทอดของวิทยากร การตอบข้อซักถามของวิทยากร ความเหมาะสม เวลา เอกสารประกอบ - ด้านการบริหารจัดการอบรม ประเด็นรูปแบบการจัดอบรม ระยะเวลา การประสานงาน	ผู้อบรม	แบบสอบถาม ความพึงพอใจ	ค่าเฉลี่ย และ S.D. เกณฑ์ตัดสิน $\bar{X} \geq 3.51$
	2. ด้านเรียนรู้ (learning)	ความรู้และเข้าใจเรื่องข้อสอบปรนัย	- มีความรู้ในหลักการพื้นฐานของข้อสอบปรนัย การวางแผนและจัดทำตารางกำหนดเนื้อหาข้อสอบ การสร้างโจทย์และตัวเลือกข้อสอบปรนัย ข้อผิดพลาดที่พบบ่อย การพิจารณาข้อสอบ การผลิตและการจัดสอบ การตรวจและวิเคราะห์ข้อสอบปรนัย การจัดทำคลังข้อสอบ การตั้งเกณฑ์ผ่านการสอบ การตัดเกรด		แบบทดสอบ	การวิเคราะห์ผลการเรียนรู้ จากคะแนนก่อนและหลังอบรม โดยสถิติ t-test dependent เกณฑ์ตัดสิน หลังการอบรมมีคะแนนความรู้สูงขึ้นอย่างมีนัยสำคัญ 0.05
ระยะที่ 2 การประเมินหลังจัดอบรมแล้ว 3 เดือน	3. ด้านพฤติกรรม (behavior)	พฤติกรรมของผู้อบรมในการนำความรู้ที่ได้ไปใช้พัฒนาข้อสอบปรนัย	ทักษะการออกข้อสอบหลังอบรม	ผู้อบรม และผู้บังคับบัญชา	แบบสอบถาม เพื่อประเมินทักษะการพัฒนาข้อสอบปรนัย	ค่าเฉลี่ย และ S.D. เกณฑ์ตัดสิน $\bar{X} \geq 3.51$
	4. ด้านผลลัพธ์ (results)	ประโยชน์ที่เกิดขึ้นต่อคณะฯ มหาวิทยาลัย อาจารย์ และนักศึกษา	การเปลี่ยนแปลงพฤติกรรมผู้อบรม		แบบสอบถาม สำหรับประเมินผลลัพธ์ที่เกิดขึ้นหลังจากอบรม	ค่าเฉลี่ย และ S.D. เกณฑ์ตัดสิน $\bar{X} \geq 3.51$

ผลการวิจัย

กลุ่มตัวอย่างที่ตอบรับเข้าร่วมวิจัย 27 คน คิดเป็นร้อยละ 75 ของประชากร ผลการประเมินประสิทธิผลหลักสูตรอบรมฯ แต่ละด้าน ดังนี้

1. ระดับปฏิกิริยา (reaction) ตามความคิดเห็นของผู้อบรม 25 คน คิดเป็นร้อยละ 92.59 ของผู้ตอบรับเข้าร่วมวิจัย ดังตารางที่ 2

ตารางที่ 2 ผลการประเมินระดับภูิกิริยา

(N = 25)

ประเด็นประเมิน	ค่าเฉลี่ย	S.D.	ระดับความพึงพอใจ	ผลประเมิน
ด้านหัวข้อการจัดอบรม				
1. ความรู้ เนื้อหา สาร	4.78	0.43	มากที่สุด	ผ่าน
2. การประยุกต์ใช้ให้เกิดประโยชน์	4.72	0.47	มากที่สุด	ผ่าน
3. เทคนิคการถ่ายทอดของวิทยากร	4.70	0.47	มากที่สุด	ผ่าน
4. การตอบข้อซักถามของวิทยากร	4.72	0.45	มากที่สุด	ผ่าน
5. ความเหมาะสมของเวลาในแต่ละหัวข้อ	4.61	0.54	มากที่สุด	ผ่าน
6. ความเหมาะสมของเอกสารประกอบ	4.60	0.60	มากที่สุด	ผ่าน
รวม	4.64	0.60	มากที่สุด	ผ่าน
ด้านการบริหารจัดการหลักสูตรอบรมฯ				
1. รูปแบบการจัดอบรม	4.68	0.48	มากที่สุด	ผ่าน
2. ระยะเวลาจัดอบรม	4.56	0.65	มากที่สุด	ผ่าน
3. การประสานงาน	4.44	0.87	มาก	ผ่าน
รวม	4.56	0.68	มากที่สุด	ผ่าน

จากตารางที่ 2 เห็นได้ว่า ระดับความพึงพอใจต่อด้านหัวข้อการอบรมโดยรวมอยู่ในระดับมากที่สุด ($\bar{X} = 4.64$, S.D. = 0.60) หากพิจารณาเป็นรายหัวข้อ พบว่า ทุกหัวข้อมีระดับความพึงพอใจมากที่สุด ($\bar{X} = 4.60$ - 4.78, S.D. = 0.43 - 0.60) ประเด็นมากที่สุด คือ ความรู้ เนื้อหา สารประโยชน์ ($\bar{X} = 4.78$, S.D. = 0.43) สำหรับความพึงพอใจต่อด้านการบริหารจัดการหลักสูตรอบรมฯ โดยรวมอยู่ในระดับมากที่สุด ($\bar{X} = 4.56$, S.D. = 0.68) หากพิจารณาเป็นรายหัวข้อ พบว่า รูปแบบการจัดอบรมอยู่ในระดับมากที่สุด ($\bar{X} = 4.68$, SD = 0.48) เมื่อเปรียบเทียบตามเกณฑ์ตัดสิน พบว่า สูงกว่าเกณฑ์กำหนด

2. ระดับเรียนรู้ (learning) ประเมินโดยวัดคะแนนความรู้ก่อนและหลังอบรมของผู้อบรม จากคะแนนทดสอบของผู้ตอบรับเข้าร่วมการอบรม 20 คน (ร้อยละ 74.07) มีผลประเมินด้านการเรียนรู้ ดังตารางที่ 3

ตารางที่ 3 เปรียบเทียบคะแนนวัดความรู้ก่อนอบรมและหลังจากอบรม

คะแนน	n	\bar{X}	S.D.	Paired differences		t	df	Sig.
				\bar{X}	S.D.			
ก่อนอบรม	20	13.60	3.30	8.50	3.41	11.15*	19	0.000
หลังอบรม	20	22.10	3.21					

* < 0.05

จากตารางที่ 3 แสดงผลวิเคราะห์การเรียนรู้ของผู้อบรม ซึ่งมีคะแนนเฉลี่ยก่อนอบรม 13.60 และคะแนนเฉลี่ยหลังอบรม 22.10 จะเห็นว่า หลังจากอบรมมีคะแนนความรู้สูงขึ้นอย่างมีนัยสำคัญทางสถิติ 0.05 โดยผ่านตามเกณฑ์กำหนด

3. ระดับพฤติกรรม (behavior) ประเมินพฤติกรรมของผู้อบรมในการนำความรู้ที่ได้ใช้พัฒนาข้อสอบปรนัย โดยการสอบถามพฤติกรรมที่ทำหลังจากอบรม เพื่อประเมินระดับทักษะการปฏิบัติ ตามความเห็นผู้อบรม 12 คน (ร้อยละ 44.44) และผู้บังคับบัญชา 10 คน (ร้อยละ 37.04) ดังตารางที่ 4

ตารางที่ 4 ผลการวิเคราะห์ทักษะการออกข้อสอบหลังอบรม

ทักษะการออกข้อสอบหลังอบรม (ตามหัวข้ออบรม)	ผู้อบรม (n = 12)				ผู้บังคับบัญชา (n = 10)			
	\bar{X}	S.D.	ระดับ	ผล	\bar{X}	S.D.	ระดับ	ผล
1. การพัฒนาข้อสอบปรนัย	3.83	1.47	บ่อยครั้ง	ผ่าน	4.89	0.33	เป็นประจำ	ผ่าน
2. การวางแผนและจัดทำตารางกำหนดเนื้อหาข้อสอบ	4.10	1.29	บ่อยครั้ง	ผ่าน	4.60	0.52	เป็นประจำ	ผ่าน
3. การตรวจสอบว่าโจทย์ข้อสอบปรนัยมีความชัดเจน	4.08	1.51	บ่อยครั้ง	ผ่าน	4.60	0.52	เป็นประจำ	ผ่าน
4. มีส่วนร่วมการผลิตข้อสอบหรือการจัดสอบ	3.92	1.44	บ่อยครั้ง	ผ่าน	4.60	0.52	เป็นประจำ	ผ่าน
5. การพิจารณาข้อสอบ	4.17	1.40	บ่อยครั้ง	ผ่าน	4.60	0.52	เป็นประจำ	ผ่าน
6. การใช้สถิติในการวิเคราะห์ข้อสอบ	4.08	1.24	บ่อยครั้ง	ผ่าน	5.00	0.00	เป็นประจำ	ผ่าน
7. การจัดทำคลังข้อสอบ	3.73	1.35	บ่อยครั้ง	ผ่าน	3.00	1.80	บางครั้ง	ไม่ผ่าน
8. การเลือกวิธีสำหรับใช้ตั้งเกณฑ์ผ่าน	3.91	1.38	บ่อยครั้ง	ผ่าน	4.60	0.52	เป็นประจำ	ผ่าน
9. การตัดเกรดให้กับนักศึกษา	3.30	1.70	บางครั้ง	ไม่ผ่าน	4.60	0.52	เป็นประจำ	ผ่าน
รวม	3.91	1.39	บ่อยครั้ง	ผ่าน	4.51	0.87	เป็นประจำ	ผ่าน

จากตารางที่ 4 แสดงว่า การประเมินผลพฤติกรรมของผู้อบรมในการนำความรู้ที่ได้ใช้พัฒนาข้อสอบปรนัย โดยรวมปฏิบัติในระดับบ่อยครั้ง ($\bar{X} = 3.91$, S.D. = 1.39) เมื่อพิจารณาเป็นรายหัวข้อ พบว่า หัวข้อที่นำไปปฏิบัติมากที่สุด คือ การพิจารณาข้อสอบ ($\bar{X} = 4.17$, S.D. = 1.40) หัวข้อที่ปฏิบัติน้อยที่สุด คือ การตัดเกรดให้กับนักศึกษา ($\bar{X} = 3.30$, S.D. = 1.70) ซึ่งปฏิบัติเป็นบางครั้ง สำหรับผู้บังคับบัญชาเห็นว่า โดยรวมปฏิบัติเป็นประจำ เมื่อพิจารณารายข้อพบว่า หัวข้อที่มีผู้บังคับบัญชาเห็นว่าปฏิบัติมากที่สุด คือ การใช้สถิติในการวิเคราะห์ข้อสอบ ($\bar{X} = 5.00$, S.D. = 0.00) หัวข้อที่ปฏิบัติน้อยที่สุด คือ การจัดทำคลังข้อสอบ ปฏิบัติเป็นบางครั้ง ($\bar{X} = 3.00$, S.D. = 1.80) เมื่อเปรียบเทียบเกณฑ์ตัดสิน พบว่า ภาพรวมคะแนนสูงกว่าเกณฑ์กำหนด

4. ระดับผลลัพธ์ (results) ประโยชน์ที่เกิดขึ้นกับสถาบัน อาจารย์ นักศึกษา ผู้วิจัยประเมินผลลัพธ์หลังจากอบรม ตามความเห็นผู้อบรมและผู้บังคับบัญชา ดังตารางที่ 5

ตารางที่ 5 การประเมินผลลัพธ์ที่เกิดขึ้นกับสถาบัน อาจารย์ นักศึกษา

ด้านผลลัพธ์ (results) ส่งผลต่อ	ผู้อบรม (n = 12)				ผู้บังคับบัญชา (n = 10)			
	\bar{X}	S.D.	ระดับ	ผล	\bar{X}	S.D.	ระดับ	ผล
สถาบัน								
1. มีการจัดการศึกษาที่มีประสิทธิภาพเพิ่มขึ้น	4.50	0.67	มากที่สุด	ผ่าน	4.60	0.52	มากที่สุด	ผ่าน
2. มีบุคลากรที่มีศักยภาพในการวัดและประเมินผลเพิ่มขึ้น	4.58	0.51	มากที่สุด	ผ่าน	4.60	0.52	มากที่สุด	ผ่าน
3. มีการจัดสอบเป็นมาตรฐานเป็นที่ยอมรับ	4.50	0.52	มากที่สุด	ผ่าน	4.60	0.52	มากที่สุด	ผ่าน
อาจารย์								
1. นำความรู้ไปปรับปรุงพัฒนาข้อสอบปรนัย	4.75	0.45	มากที่สุด	ผ่าน	4.60	0.52	มากที่สุด	ผ่าน
2. เผยแพร่ความรู้จากการอบรมให้กับเพื่อนอาจารย์ และผู้ที่เกี่ยวข้องในสถาบัน	3.92	0.79	มาก	ผ่าน	3.80	1.55	มาก	ผ่าน

ตารางที่ 5 การประเมินผลลัพธ์ที่เกิดขึ้นกับสถาบัน อาจารย์ นักศึกษา (ต่อ)

ด้านผลลัพธ์ (results) ส่งผลดีต่อ	ผู้อบรม (n = 12)				ผู้บังคับบัญชา (n = 10)			
	\bar{X}	S.D.	ระดับ	ผล	\bar{X}	S.D.	ระดับ	ผล
3. มีความมั่นใจในการออกข้อสอบมากขึ้น	4.67	0.49	มากที่สุด	ผ่าน	4.10	0.32	มาก	ผ่าน
4. นำความรู้ไปปรับปรุงพัฒนาการจัดสอบ	4.58	0.51	มากที่สุด	ผ่าน	4.60	0.52	มากที่สุด	ผ่าน
นักศึกษา								
1. ผู้เรียนได้รับความเป็นธรรมในการสอบ	4.67	0.49	มากที่สุด	ผ่าน	4.60	0.52	มากที่สุด	ผ่าน
2. ผู้เรียนมีความมั่นใจในการจัดสอบ	4.42	0.78	มาก	ผ่าน	4.20	1.03	มาก	ผ่าน
3. ไม่มีข้อโต้แย้งกับผลการสอบ	4.33	0.78	มาก	ผ่าน	5.00	0.00	มากที่สุด	ผ่าน
รวม	4.49	0.62	มาก	ผ่าน	4.46	0.78	มาก	ผ่าน

จากตารางที่ 5 แสดงว่า โดยรวมผู้อบรมเห็นด้วยกับผลลัพธ์ของหลักสูตรอบรมฯ อยู่ในระดับมาก ($\bar{X} = 4.49$, S.D. = 0.62) เมื่อพิจารณาเป็นรายประเด็น พบว่า ผลลัพธ์ที่เกิดขึ้นส่งผลดีต่ออาจารย์ในระดับมากที่สุด ($\bar{X} = 4.75$, S.D. = 0.45) คือ การนำประเด็นนำความรู้ไปปรับปรุงพัฒนาข้อสอบปรนัย สำหรับผู้บังคับบัญชา เห็นด้วยกับผลลัพธ์โดยรวมในระดับมาก ($\bar{X} = 4.46$, S.D. = 0.78) เมื่อพิจารณาเป็นรายประเด็น พบว่า ผลลัพธ์ที่เกิดขึ้นส่งผลดีต่อนักศึกษามากที่สุด ($\bar{X} = 5.00$, S.D. = 0.00) คือ นักศึกษาไม่มีข้อโต้แย้งกับผลการสอบ เมื่อเปรียบเทียบเกณฑ์ตัดสินพบว่า ภาพรวมคะแนนสูงกว่าเกณฑ์กำหนด

อภิปรายผล

1. การประเมินปฏิกิริยา (reaction) ของหลักสูตรฯ พบว่า คะแนนเฉลี่ยภาพรวมและทุกด้านผ่านเกณฑ์ที่กำหนด โดยความพึงพอใจต่อหัวข้อการอบรมทุกหัวข้อมากที่สุด และจากข้อคำถามปลายเปิดแสดงให้เห็นว่า วิทยากรที่ถ่ายทอดความรู้ เนื้อหา สาระ ในแต่ละหัวข้อมีความเชี่ยวชาญ และมีประสบการณ์ โดยสามารถถ่ายทอดความรู้ได้ตรงประเด็น มีการใช้เทคนิคการถ่ายทอดที่เหมาะสม อาทิ ให้ผู้เข้าอบรมร่วมตอบและแสดงความคิดเห็น การแบ่งกลุ่ม มียกตัวอย่าง มอบหมายงานเพื่อเพิ่มความเข้าใจ เปิดโอกาสให้ซักถาม ทำให้รู้สึกว่าการเกิดประโยชน์สามารถประยุกต์ใช้ความรู้ โดยมีเอกสารประกอบการอบรม และใช้เวลาในแต่ละหัวข้อได้อย่างเหมาะสม ดังที่ ซูซึ สมิตทิกไร (Smithikrai, 2001) กล่าวว่า วิธีการฝึกอบรมช่วยให้เกิดการเรียนรู้ หรือ การเปลี่ยนแปลงด้านความรู้ คือ การบรรยาย ส่วนวิธีฝึกอบรมที่ช่วยเปลี่ยนแปลงพฤติกรรม ทักษะ หรือ ความสามารถ ได้แก่ การฝึกปฏิบัติ กรณีศึกษา การสาธิต การจำลองสถานการณ์ การถ่ายทอดงานหรือการสนทนากัน

ในส่วนของความพึงพอใจต่อการบริหารจัดการหลักสูตรอบรมฯ ประเด็นรูปแบบและระยะเวลาจัดอบรม มีค่าเฉลี่ยมากที่สุด นั้นแสดงว่า การจัดอบรมรูปแบบออนไลน์และระยะเวลาอบรม 2 วัน มีความเหมาะสม ส่วนประเด็นการประสานงาน ระดับมาก แสดงว่า การจัดอบรมรูปแบบออนไลน์ที่ศูนย์ฯ ต้องปรับตามสถานการณ์โควิด-19 ผู้อบรมรู้สึกว่าจะเหมาะสมดีแล้ว แต่เมื่อมีการปรับเปลี่ยนอย่างรวดเร็วจึงทำให้มีความคาดหวังเรื่องการประสานงานจากเจ้าหน้าที่เพื่ออำนวยความสะดวกมากกว่าปกติ สอดคล้องกับที่ดุสิต ชาวเหลียง (Khawlueng, 2011) กล่าวว่า การพัฒนาโปรแกรมฝึกอบรมมีรายละเอียดเกี่ยวกับทุกสิ่งทุกอย่างของการฝึกอบรมที่จะต้อง

ปฏิบัติ ซึ่งมีความสำคัญไม่ว่าจะเป็นสื่อการเรียนการสอน เครื่องมือ อุปกรณ์ที่ใช้ การระบุเนื้อหาที่จำเป็นต่อการเรียนรู้ การเลือกสิ่งอำนวยความสะดวกต่อการอบรม

2. การประเมินเรียนรู้ (learning) พบว่า หลังจากอบรมมีคะแนนความรู้สูงขึ้น อย่างมีนัยสำคัญทางสถิติ 0.05 โดยผ่านตามเกณฑ์ที่กำหนด แสดงว่า การจัดอบรมที่มีการใช้รูปแบบ เทคนิคการถ่ายทอดของวิทยากรที่เหมาะสมและมีประสิทธิภาพจะทำให้เกิดการเรียนรู้ที่ดีที่สุด ดังที่ พยอม วงศ์สารศรี (Wongsarnsri, 1998) กล่าวว่า การฝึกอบรมนั้นถือว่าเป็นวิธีการหรือกระบวนการที่ทำให้บุคคลในฐานะองค์ประกอบขององค์กรได้เกิดการเรียนรู้ และจะช่วยให้ปฏิบัติงานสำเร็จลุล่วงด้วยดี

3. การประเมินพฤติกรรม (behavior) พบว่า ตามความเห็นของผู้อบรม มีการนำความรู้ที่ได้ไปใช้พัฒนาข้อสอบปรนัย โดยรวมปฏิบัติบ่อยครั้งและสูงกว่าเกณฑ์กำหนด หัวข้อที่ปฏิบัติมากที่สุด คือ การพิจารณาข้อสอบ การวางแผนและจัดทำตารางกำหนดเนื้อหาข้อสอบ การตรวจสอบว่าโจทย์ข้อสอบปรนัยมีความชัดเจนและการใช้สถิติในการวิเคราะห์ข้อสอบ ซึ่งทิศทางเดียวกับความเห็นผู้บังคับบัญชา สำหรับหัวข้อที่นำไปปฏิบัติน้อยที่สุดและไม่ผ่านตามเกณฑ์กำหนด คือ การจัดทำคลังข้อสอบ แสดงว่า ผู้บังคับบัญชายังไม่เห็นการจัดทำคลังข้อสอบที่เป็นรูปธรรมมากนักและสำหรับความเห็นของผู้อบรมเรื่องการตัดเกรดให้กับนักศึกษา ($\bar{X} = 3.30$ ระดับบางครั้ง) ซึ่งมีคะแนนเฉลี่ยและระดับแตกต่างกับความเห็นของผู้บังคับบัญชา ($\bar{X} = 4.60$ ระดับเป็นประจำ) นั้นอาจเพราะผู้อบรมคิดว่ยังไม่ค่อยมีโอกาสตัดเกรดบ่อยนัก แต่ผู้บังคับบัญชาอาจมองว่า ความรู้นี้สามารถที่จะนำไปใช้ประโยชน์ได้ ซึ่งส่วนหนึ่งอาจเกิดจากข้อจำกัดการวิจัย เนื่องจากผู้อบรมอาจยังไม่มีโอกาสตัดเกรดมากพอดังนั้น สถาบันอาจมีการพัฒนาระบบคลังข้อสอบและระบบการตัดเกรด เพื่อช่วยส่งเสริม ผลักดันให้เกิดการปฏิบัติและการนำไปใช้ได้มากขึ้นด้วย เช่น การสัมมนาเพื่อพิจารณาข้อสอบและการจัดเก็บในคลังข้อสอบ ดังที่ คณะแพทยศาสตร์ศิริราชพยาบาลจัดให้มีการสัมมนาระหว่างส่วนงานเรื่องการตัดเกรด เพื่อแลกเปลี่ยนแนวคิดและประสบการณ์ในการตัดเกรดซึ่งจะทำให้มีโอกาสอภิปรายกันอย่างกว้างขวางถึงวิธีการ ปัญหาที่พบเพื่อนำไปใช้และปรับปรุงวิธีการตัดเกรด (Iramaneerat, 2008)

4. การประเมินผลลัพธ์ (results) ผู้อบรมและผู้บังคับบัญชามีความเห็นเป็นทิศทางเดียวกัน คือ เห็นด้วยกับผลลัพธ์ของหลักสูตรฯ โดยรวมเกิดประโยชน์ต่อสถาบัน อาจารย์ และนักศึกษาขึ้นอย่างมาก เมื่อพิจารณารายประเด็นเห็นว่า ผู้อบรมเห็นด้วยว่าผลลัพธ์ที่เกิดขึ้นนั้นส่งผลที่ดีกับอาจารย์ในประเด็นนำความรู้ไปปรับปรุงพัฒนาข้อสอบปรนัย รองลงมาเป็นผลลัพธ์ที่เกิดขึ้นส่งผลต่อนักศึกษาในประเด็นผู้เรียนจะได้รับความเป็นธรรมในการสอบ สำหรับความเห็นผู้บังคับบัญชาคิดว่า ผลลัพธ์ที่เกิดขึ้นส่งผลต่อนักศึกษาในประเด็นไม่มีข้อโต้แย้งกับผลการสอบมากที่สุด สำหรับค่าเฉลี่ยของผลลัพธ์ที่ส่งผลต่ออาจารย์ในประเด็นเผยแพร่ความรู้จากการอบรมให้กับเพื่อนอาจารย์และผู้มีส่วนเกี่ยวข้องในสถาบัน น้อยที่สุดแต่ยังอยู่ในเกณฑ์ผ่าน นั้นแสดงว่า ยังไม่มีโอกาสถ่ายทอดความรู้จากการอบรม หรือแลกเปลี่ยนเรียนรู้ให้คนอื่น ดังนั้น ควรมีช่องทางหรือเวทีให้นำความรู้มาเผยแพร่ เช่น กำหนดในวาระการประชุม ฯลฯ จนเกิดเป็นผลลัพธ์ที่ส่งผลต่อองค์กร สอดคล้องกับองค์ประกอบสำคัญที่ส่งผลต่อพฤติกรรมของบุคลากร ซึ่งการแลกเปลี่ยนความรู้ระหว่างกันจะช่วยทำให้เกิดการเรียนรู้และแนวคิดใหม่ ๆ ในการพัฒนาองค์กร (Alam, Abdullah, Ishak, & Zain, 2009)

สรุปผล

หลักสูตรอบรมการพัฒนาข้อสอบปรนัยเพื่อประเมินความรู้ทางการแพทย์ ของศูนย์ฯ เมื่อเปรียบเทียบเกณฑ์ตัดสิน พบว่า คะแนนโดยรวมทุกด้านเป็นตามเกณฑ์ที่กำหนด โดยแต่ละด้านการประเมินสรุปได้ดังนี้

1. ปฏิกริยา (reaction) ความเหมาะสมของหัวข้อและการบริหารจัดการหลักสูตรอบรมฯ อยู่ในระดับดีมาก
2. เรียนรู้ (learning) มีความรู้และเข้าใจ เรื่อง MCQ มากขึ้นกว่าก่อนอบรม
3. พฤติกรรม (behavior) หลักสูตรอบรมฯ ทำให้มีพฤติกรรมที่พัฒนาทักษะการออกข้อสอบ โดยผู้อบรมมีการปฏิบัติระดับบ่อยครั้ง และผู้บังคับบัญชาเห็นว่า ปฏิบัติอยู่เป็นประจำ
4. ผลลัพธ์ (results) หลักสูตรอบรมฯ มีผลลัพธ์ที่ส่งผลดีต่อสถาบัน อาจารย์ และนักศึกษา โดยผู้อบรมและผู้บังคับบัญชาเห็นด้วยในระดับมาก

ข้อเสนอแนะ

ข้อเสนอแนะจากการวิจัย

1. จากผลการวิจัย พบว่า การนำไปปฏิบัติที่มีค่าเฉลี่ยน้อยที่สุดและไม่ผ่านตามเกณฑ์กำหนด คือ การจัดทำคลังข้อสอบ ดังนั้น สถาบันควรให้ความรู้เรื่องการจัดทำคลังข้อสอบมากยิ่งขึ้น หรือมีข้อกำหนดในการจัดทำคลังข้อสอบเพื่อกระตุ้นและส่งเสริมให้เกิดการนำไปใช้
2. จากผลวิจัยเห็นได้ว่า ผลลัพธ์ในการเผยแพร่ความรู้จากการอบรมให้กับเพื่อนอาจารย์และผู้ที่เกี่ยวข้องในสถาบันมีค่าเฉลี่ยน้อยที่สุด ดังนั้น การเพิ่มประสิทธิภาพของการอบรมให้เกิดผลลัพธ์กับสถาบัน อาจารย์ และนักศึกษานั้น จำเป็นต้องมีนโยบายหรือการส่งเสริมให้เกิดการปฏิบัติและการนำไปใช้มากขึ้น

ข้อเสนอแนะในการวิจัยครั้งถัดไป

1. ควรออกแบบหรือวางแผนการติดตามผลการอบรม โดยเน้นกระบวนการติดตามเพื่อให้เห็นการเปลี่ยนแปลง เช่น การวิจัยเชิงทดลอง เปรียบเทียบประเมินคุณภาพข้อสอบ
2. ควรติดตามผลลัพธ์จากการอบรมที่เกิดกับสถาบัน อาจารย์ และนักศึกษาในระยะยาว และต่อเนื่อง เพื่อให้เห็นผลลัพธ์ชัดเจนยิ่งขึ้น

ข้อจำกัดของการวิจัย

กลุ่มตัวอย่างเป็นทั้งบุคคลภายนอกและภายในคณะแพทยศาสตร์ศิริราชพยาบาล มหาวิทยาลัยมหิดล จึงเกิดความแตกต่างตามบริบท และระยะเวลาเก็บข้อมูลหลังสิ้นสุดอบรม 3 เดือน อาจทำให้ยังไม่เห็นผลลัพธ์เท่าที่ควร และอัตราการตอบกลับน้อย จึงอาจส่งผลให้เกิดความคลาดเคลื่อนของข้อมูล

จริยธรรมการวิจัยในมนุษย์

รับรองโดยสำนักงานจริยธรรมการวิจัยในคน (certificate of approval) คณะแพทยศาสตร์ศิริราชพยาบาล มหาวิทยาลัยมหิดล

กิตติกรรมประกาศ

งานวิจัยนี้สำเร็จได้ด้วยความกรุณาจากผู้เชี่ยวชาญทุกท่านที่กรุณาสละเวลาอันมีค่ายิ่งเพื่อตรวจสอบและให้ข้อเสนอแนะ และขอขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร.วรรณ วาณิชย์เจริญชัย ซึ่งเป็นที่ปรึกษา งานวิจัย โดยงานวิจัยนี้ได้รับทุนสนับสนุนจากกองทุนพัฒนาแพทยศาสตร์ คณะแพทยศาสตร์ศิริราชพยาบาล ปี พ.ศ. 2564

References

- Alam, S. S., Abdullah, Z., Ishak, N. A., & Zain, Z. M. (2009). Assessing Knowledge Sharing Behaviour Among Employees in SMEs: An Empirical Study. **International Business Research**, 2(2): 115-122.
- Cohen, J. (1977). **Statistical Power Analysis for the Behavioral Sciences**. New York: Academic Press.
- Iramaneerat, Cherdsak. (2008). Grading (การตัดเกรด). **Siriraj Medical Bulletin**, 1(2): 107-113.
- Iramaneerat, Cherdsak. (2012). Creating a Multiple-Choice Test (การสร้างข้อสอบปรนัย). **Siriraj Medical Bulletin**, 5(1): 29-36.
- Khawloueng, Dusit. (2011). Competency and Effective Training (การฝึกอบรมที่มีประสิทธิภาพและสมรรถนะ). **Journal of Education and Social Development**, 7(1): 18-32.
- Kirkpatrick, D. L. (1994). **Evaluating Training Programs: The Four Levels**. San Francisco: Berrett-Koehler Publishers.
- Panakul, Saksri. (2002). **Curriculum Evaluation (การประเมินหลักสูตร) (3rd ed.)**. Bangkok: Ramkhamhaeng University Press.
- Silpcharu, Thanin. (2006). **Research and Statistical Analysis with SPSS (การวิจัยและการวิเคราะห์ข้อมูลทางสถิติด้วย SPSS) (5th ed.)**. Bangkok: V. Inter Print.
- Smithikrai, Chuchai. (2001). **Personnel Training in Organizations (การฝึกอบรมบุคลากรในองค์การ) (3rd ed.)**. Bangkok: Chulalongkorn University Press.
- Wongsarnsri, Payom. (1998). **Human Resource Management (การบริหารทรัพยากรมนุษย์) (7th ed.)**. Bangkok: Suan Dusit Rajabhat Institute.