

# จุดมุ่งหมายทางการศึกษาด้านพุทธิปริเขต ดัชนีความยากง่าย ดัชนีอำนาจจำแนกและประสิทธิภาพตัวลองของข้อสอบ แบบเลือกตอบ กลุ่มวิชาชีพหลักสูตรพยาบาลศาสตรบัณฑิต วิทยาลัยพยาบาลตำรวจ ปีการศึกษา 2556\*

พ.ต.อ.หญิง ดร. กัญญาภัทรา เกษโกมล\*\*

พ.ต.อ.หญิง ดร. ปราณีย์ เสนีย์\*\*

## บทคัดย่อ

การศึกษาครั้งนี้มีวัตถุประสงค์เพื่อวิเคราะห์จุดมุ่งหมายทางการศึกษาด้านพุทธิปริเขต ระดับความยากง่ายและอำนาจจำแนก ประสิทธิภาพตัวลองของข้อสอบ และวิเคราะห์ความสัมพันธ์ระหว่างค่าความยากง่ายและอำนาจจำแนกของข้อสอบแบบเลือกตอบ กลุ่มวิชาชีพ หลักสูตรพยาบาลศาสตรบัณฑิต วิทยาลัยพยาบาลตำรวจ ปีการศึกษา 2556 โดยศึกษาข้อมูลจากข้อสอบ และผลการวิเคราะห์ข้อสอบของข้อสอบปลายภาค กลุ่มวิชาชีพ ปีการศึกษา 2556 จำนวน 19 รายวิชา รวมข้อสอบ 19 ฉบับ วิเคราะห์ข้อมูลพื้นฐานของข้อสอบ และความสัมพันธ์ระหว่างดัชนีความยากง่ายและอำนาจจำแนกของข้อสอบ และนำเสนอข้อมูลในรูปแบบตาราง กราฟ และแผนภาพกระจาย

ผลการศึกษาพบว่า ระดับการวัดตามจุดมุ่งหมายทางการศึกษาด้านพุทธิปริเขต มีเพียง 5 ระดับ ได้แก่ ความรู้ความจำ ความเข้าใจ การนำไปใช้ การวิเคราะห์ และการประเมินค่า ซึ่งมีเพียง 2 ฉบับเท่านั้น ที่เน้นการวัดความคิดระดับสูง ค่าเฉลี่ยของดัชนีความยากง่ายของข้อสอบทั้ง 19 ฉบับ อยู่ในเกณฑ์ ปานกลาง (0.45) ถึง ค่อนข้างง่าย (0.77) ส่วนค่าเฉลี่ยของอำนาจจำแนกของข้อสอบทั้ง 19 ฉบับ อยู่ในเกณฑ์ต่ำมาก ควรปรับปรุง (0.08) ถึง พอใช้ได้ (0.31) ตัวลองที่ดีทั้ง 3 ตัวเลือก พบเพียง 1-3 ข้อ คิดเป็นร้อยละ 1.67-10.00 ใน 10 รายวิชา ไม่พบความสัมพันธ์เชิงเส้นตรงระหว่างดัชนีความยากง่ายและอำนาจจำแนกของข้อสอบ จำนวน 16 ฉบับจาก 19 ฉบับ และทิศทางความสัมพันธ์ระหว่างดัชนีความยากง่ายและอำนาจจำแนกของข้อสอบเป็นบวกในจำนวน 15 ฉบับ จาก 19 ฉบับ

**คำสำคัญ :** จุดมุ่งหมายทางการศึกษาด้านพุทธิปริเขต ดัชนีความยากง่าย ดัชนีอำนาจจำแนก และประสิทธิภาพตัวลอง

\* ได้รับทุนอุดหนุนจากวิทยาลัยพยาบาลตำรวจ

\*\* ผู้ช่วยศาสตราจารย์ (สบ 4) กลุ่มงานอาจารย์ วิทยาลัยพยาบาลตำรวจ

# The Cognitive Domain of Educational Objectives, Item Difficulty Indices, Item Discrimination Indices and Distracter Analysis in Multiple Choice Questions in Professional Education Courses, Bachelor of Nursing Science Program, Police Nursing College, Academic Year 2013

Assistant Professor Police Colonel Thipkhumporn Keskomon, Ph.D.

Assistant Professor Police Colonel Pranee Sanee, Ph.D.

## Abstract

The purposes of this study were to analyze the cognitive domain of educational objectives, item difficulty indices, item discrimination indices and distracters and to examine the relationship between item difficulty indices and item discrimination indices in multiple choice questions in Professional Education Courses, Bachelor of Nursing Science Program, Police Nursing College in Academic Year 2013. Nineteen tests of nineteen subjects were used for the study. Descriptive statistics and Pearson Product Moment correlation were used for analyzing data. Data were presented in tables, graphs and scatter diagrams.

The results showed that: There were five types of the cognitive domain of educational objective: knowledge, understanding, application, analysis and evaluation. There were only two tests focusing on higher thinking levels. The average of the item difficulty indices of 19 tests was moderate (0.45) to relatively easy (0.77) and the average of the item discrimination indices of the 19 tests remained very low (0.08) to fair (0.31). Three efficient distracters were found only 1-3 items (1.67%-10.00%) in 10 tests. No significant linear relationships between the item difficulty indices and item discrimination indices were found in 16 of 19 tests. And the relationships between the item difficulty indices and item discrimination indices were positive in 15 of 19 tests.

**Keywords :** Cognitive Domain of Educational Objectives, Item Difficulty Indices, Item Discrimination Indices, Distracter

## ความเป็นมาและความสำคัญของปัญหา

ข้อสอบถือเป็นเครื่องมือสำคัญในการวัดผลการเรียนรู้ของผู้เรียน ปัจจุบันพบว่าข้อสอบสามารถใช้ในการพัฒนาผู้เรียนได้ ทั้งในเนื้อหาวิชาและทักษะทางปัญญา ซึ่งเป็นจามิน เอส บลูม และคณะ (Bloom et al., 1956 อ้างใน Reilly & Oermann, 1999) ได้แบ่งวัตถุประสงค์ทางการศึกษาแบ่งออกเป็น 3 โดเมน ได้แก่ 1) พุทธิปริเขต (Cognitive Domain) 2) จิตตปริเขต (Affective Domain) และ 3) ทักษะปริเขต (Psychomotor Domain) พุทธิปริเขต แบ่งเป็น 6 ชั้น ได้แก่ ความรู้ ความเข้าใจ การประยุกต์/การนำไปใช้ การวิเคราะห์ การสังเคราะห์ การประเมินค่า ข้อสอบแบบเลือกตอบนิยมใช้ในการวัดทางพุทธิปริเขตมากที่สุด

ข้อสอบแบบเลือกตอบ ได้รับความนิยมจากสถาบันการศึกษาต่าง ๆ และให้สารสนเทศ/ข้อมูลป้อนกลับแก่ผู้สอนเกี่ยวกับความสามารถในการสอนของผู้สอน และความสามารถในการเรียนรู้ของผู้เรียน การเรียนการสอนทางการแพทย์ ได้นำข้อสอบแบบเลือกตอบมาใช้มากที่สุด เพื่อเป็นเครื่องมือในการประเมินความรู้ความสามารถของผู้เรียน (Mitra, Ponnudurai, & Judson, 2009: 2) โดยกำหนดให้ผู้สอบเลือกคำตอบที่ถูกต้องอย่างชัดเจนเพียงข้อเดียว หรือให้ผู้สอบเลือกคำตอบที่ดีที่สุดจากการเลือกคำตอบที่มีความสมเหตุสมผลหลายข้อ ข้อสอบแบบเลือกตอบที่เขียนตัวลงที่ดี จะสามารถจำแนกผู้เรียนที่เก่งออกจากผู้เรียนที่อ่อนได้

จุดแข็งของข้อสอบแบบเลือกตอบ คือสามารถวัดความรู้ความสามารถของผู้เรียนได้ทุกระดับของพุทธิปริเขต (cognitive domain) สามารถสุ่มเนื้อหาและวัตถุประสงค์ได้กว้าง ให้คะแนนที่เชื่อถือได้สูง ตรวจให้คะแนนด้วยเครื่องได้รวดเร็วและถูกต้อง โอกาสในการเดาลดลง เมื่อเปรียบเทียบกับข้อสอบถูก-ผิด อย่างไรก็ตาม การออกแบบการเขียนข้อสอบแบบเลือกตอบก็มีข้อจำกัด เพราะเป็นกระบวนการที่มีความซับซ้อน สร้างยากและใช้เวลาในการสร้างมาก (Karelia, Pillai, & Vegada, 2013: 41) ทั้งนี้ยังขึ้นอยู่กับทักษะการอ่านของผู้สอบและความสามารถของผู้เขียนข้อสอบอีกด้วย โดยข้อสอบที่วัดความรู้ในระดับต่ำเขียนได้ง่าย ทำให้ผู้สอนละเลยการเขียนข้อสอบที่วัดการคิดระดับสูง ผู้สอบสามารถเดาได้ แม้ไม่มีความรู้ที่แท้จริง อาจทำให้ผู้สอนเข้าใจผิด คิดว่าผู้เรียนมีความรู้ความสามารถ ทำให้ไม่สามารถพัฒนาผู้เรียนได้อย่างเหมาะสม

ในการออกแบบการประเมินผลเพื่อวัดผลการเรียนรู้ของผู้เรียน ผู้สอนจะต้องตอบได้ว่าผู้เรียนบรรลุผลลัพธ์การเรียนรู้ที่คาดหวังหรือบรรลุวัตถุประสงค์ทางการศึกษาที่กำหนดไว้หรือไม่ โดยสามารถตอบได้ว่าผู้เรียนรู้อะไร ทำอะไรได้ และเป็นคนประเภทใด (Karelia et al., 2013: 41) แต่ความกังวลประการหนึ่งของสถาบันการศึกษา เกี่ยวกับการสร้างข้อสอบแบบเลือกตอบ คือคุณภาพของข้อสอบ และแบบสอบ การตรวจสอบคุณภาพของข้อสอบ ทำได้โดยการวิเคราะห์ข้อสอบ ซึ่งหมายถึง

กระบวนการเก็บรวบรวมข้อมูล สรุปและใช้สารสนเทศจากการตอบของผู้เรียนเพื่อประเมินคุณภาพของข้อสอบ (Karelia et al., 2013: 41; Mitra et al., 2009: 2) ซึ่งเป็นเทคนิคของการตรวจสอบคุณภาพของข้อสอบเป็นรายข้อ

ดัชนีสำคัญที่บ่งชี้คุณภาพของข้อสอบ ตามทฤษฎีการสอบแบบดั้งเดิม (Classical test theory: CTT) ได้แก่ ความยากง่ายของข้อสอบ (Item difficulty) และอำนาจจำแนกของข้อสอบ (Item discrimination) ระดับความยากง่ายของข้อสอบ (P) หมายถึง สัดส่วนของจำนวนคนที่ตอบข้อสอบข้อนั้นถูก อำนาจจำแนกของข้อสอบ (r) หมายถึง ความสามารถของข้อสอบในการจำแนกหรือแยกความแตกต่างระหว่างผู้สอบที่มีผลสัมฤทธิ์ต่างกัน สำหรับข้อสอบแบบเลือกตอบ ผู้เขียนข้อสอบจะต้องให้ความสำคัญกับประสิทธิภาพของตัวลวงด้วย ประสิทธิภาพของตัวลวง แบ่งออกเป็น 1) สัดส่วนของผู้เลือกตัวลวง ( $P_w$ ) หมายถึง สัดส่วนของจำนวนคนที่เลือกตัวลวงนั้น ๆ และ 2) อำนาจจำแนกของตัวลวง ( $r_w$ ) หมายถึง ผลต่างระหว่างสัดส่วนของคนในกลุ่มอ่อนที่เลือกตัวลวงกับสัดส่วนของคนในกลุ่มเก่งที่เลือกตัวลวงนั้น ๆ

ความสัมพันธ์ระหว่างค่าความยากง่ายและอำนาจจำแนกของข้อสอบ พบว่าข้อสอบที่ยากเกินไปหรือง่ายเกินไปจะไม่สามารถจำแนกผู้เรียนเก่งกับอ่อนออกจากกันได้ หรือไม่สามารจำแนกผู้เรียนที่มีความสามารถพิเศษกับที่ไม่มีความสามารถพิเศษออกจากกันได้ (ศิริชัย กาญจนวาสี, 2544: 187) ผลการศึกษาเกี่ยวกับความสัมพันธ์ระหว่างดัชนีความยากง่ายและอำนาจจำแนกของข้อสอบ พบผลที่แตกต่างกัน สรุปได้ดังนี้ 1) มีความสัมพันธ์กันในระดับต่ำ อย่างไม่มีนัยสำคัญทางสถิติ (Karelia et al., 2013) 2) มีความสัมพันธ์กัน โดยพบว่าข้อสอบที่มีระดับความยากง่ายปานกลางจะมีความสามารถในการจำแนกสูงที่สุด ข้อสอบที่ง่ายเกินไปหรือยากเกินไป จะมีความสามารถในการจำแนกต่ำ (Hotiu, 2006; Pande, Pande, Parate, Nikam, & Agrekar, 2013) หรือค่าอำนาจจำแนกเป็นลบ (Sim & Rasiah, 2006) 3) ความสัมพันธ์ไม่ใช่เชิงเส้นตรง แต่เป็นรูปโดม (Sim & Rasiah, 2006) อย่างไรก็ตามข้อสอบแบบเลือกตอบที่มีอำนาจจำแนกดีมีแนวโน้มว่าข้อสอบจะมีความยากง่ายในระดับปานกลาง และข้อสอบที่มีความยากง่ายปานกลางถึงยากเกินไปมีแนวโน้มจะให้อำนาจจำแนกทางลบ

ในการศึกษาวิชาชีพพยาบาล ยังไม่พบว่า มีสถาบันการศึกษาใด ทำการวิเคราะห์จุดมุ่งหมายทางการศึกษาด้านพุทธิปริเขต และคุณภาพของข้อสอบแบบเลือกตอบทั้ง 3 พารามิเตอร์ ทำให้การนำสารสนเทศหรือข้อมูลป้อนกลับที่ได้รับไปใช้เพื่อพัฒนาคุณภาพการสอน การสอบ การจัดการเรียนรู้สำหรับผู้เรียน และการพัฒนาผู้เรียนยังไม่เกิดประโยชน์เท่าที่ควร ประกอบกับข้อสอบแบบเลือกตอบได้รับการนำมาใช้เป็นเครื่องมือในการตัดสินความรู้ความสามารถของผู้เรียน ผลที่ได้รับ

อาจไม่ถูกต้อง เพราะข้อสอบที่ผู้สอนสร้างขึ้นมีคุณภาพที่ไม่ดีพอที่จะวัดความรู้ความสามารถที่แท้จริงของผู้เรียน ดังนั้นคณะผู้วิจัยจึงสนใจที่จะวิเคราะห์จุดมุ่งหมายทางการศึกษาด้านพุทธิปริเขต ความยากง่าย อำนาจจำแนก และประสิทธิภาพตัวลงของข้อสอบแบบเลือกตอบ กลุ่มวิชาชีพ หลักสูตรพยาบาลศาสตรบัณฑิต วิทยาลัยพยาบาลตำรวจ ปีการศึกษา 2556 เพื่อนำผลที่ได้รับ มาสรุปผลการเรียนรู้ของผู้เรียนได้อย่างมั่นใจยิ่งขึ้น และผู้สอนจะสามารถนำผลที่ได้รับไปปรับปรุงคุณภาพการสอน การสอบ การจัดการเรียนรู้สำหรับผู้เรียน และการพัฒนาผู้เรียนให้บรรลุวัตถุประสงค์ในการศึกษาวิชาชีพพยาบาลต่อไป

### วัตถุประสงค์ของการวิจัย

งานวิจัยนี้มีวัตถุประสงค์เพื่อ วิเคราะห์จุดมุ่งหมายทางการศึกษาด้านพุทธิปริเขต ความยากง่ายและอำนาจจำแนก ประสิทธิภาพตัวลง และวิเคราะห์ความสัมพันธ์ระหว่างค่าความยากง่ายและอำนาจจำแนกของข้อสอบแบบเลือกตอบ กลุ่มวิชาชีพ หลักสูตรพยาบาลศาสตรบัณฑิต วิทยาลัยพยาบาลตำรวจ ปีการศึกษา 2556

### กรอบแนวคิดการวิจัยและวรรณกรรมที่เกี่ยวข้อง

สารบบจำแนกจุดมุ่งหมายทางการศึกษาด้านพุทธิปริเขต (Taxonomy of cognitive domain of educational objective) ตามที่บลูมและคณะ (Bloom et al., 1956 อ้างใน Reilly & Oermann, 1999) ได้เสนอไว้มี 6 ชั้น ดังนี้ (Reilly & Oermann, 1999; อุทุมพร ทองอุไทย, 2523)

1.00 ความรู้ (knowledge) เน้นกระบวนการจำมากที่สุด แบ่งออกเป็น 1.10 ความรู้เฉพาะสิ่ง 1.20 ความรู้เรื่องวิถีและวิธีการจัดการกระทำกับสิ่งเฉพาะ 1.30 ความรู้เรื่องสากลและเรื่องนามธรรมในสาขาต่าง ๆ

2.00 ความเข้าใจ (Comprehensive) อาศัยความสัมพันธ์กับเนื้อหาหรือเนื้อเรื่อง เพื่อสื่อความหมาย แบ่งออกเป็น 2.10 การแปล เป็นการสื่อความหมายจากรูปแบบหนึ่ง ไปยังอีกรูปแบบหนึ่ง ภายใต้เนื้อหาเดิม 2.20 การตีความ เป็นการอธิบายหรือสรุปเรื่องในการสื่อความหมาย มีการจัดเรียงลำดับ การจัดเรียงระบบ หรือให้ภาพพจน์ของเนื้อหาใหม่ 2.30 การสรุปอ้างอิง เป็นการขยายแนวโน้มหรือความโน้มเอียงเกินจากข้อมูลออกไป เพื่อสามารถกำหนดความหมายอื่น

3.00 การประยุกต์ (Application) การประยุกต์ใช้ความคิดทั่ว ๆ ไป กฎของกระบวนการ หรือวิธีการสรุป หลักเชิงเทคนิค ความคิด และทฤษฎี ซึ่งเป็นนามธรรม ในสถานการณ์ที่เป็นรูปธรรม

4.00 การวิเคราะห์ (Analysis) การจำแนกการสื่อความหมายไปสู่หน่วยย่อย ๆ หรือส่วนย่อย ๆ เพื่อให้ได้ลำดับชั้นของความคิดที่แสดงออกอย่างชัดเจน และ/หรือความสัมพันธ์ระหว่างความคิดที่แสดงออกมาให้ชัดเจน แบ่งออกเป็น 4.10 การวิเคราะห์หน่วยย่อย เป็นการชี้หน่วยย่อย ๆ ที่รวมในการสื่อความหมาย 4.20 การวิเคราะห์ความสัมพันธ์ เป็นการพิจารณาการเชื่อมโยงและปฏิสัมพันธ์ระหว่างหน่วยย่อยกับส่วนต่าง ๆ ของการสื่อความหมาย 4.30 การวิเคราะห์หลักการจัดระเบียบ เป็นการพิจารณาระบบ โครงสร้าง การจัดเรียง ในการสื่อความหมาย ทั้งที่เด่นชัดและที่เป็นนัย รวมทั้งพื้นฐานที่จำเป็น และกลไกการสื่อความหมายเป็นหน่วยเดียวกัน

5.00 การสังเคราะห์ (Synthesis) การนำหน่วยหรือส่วนต่าง ๆ มารวมเข้าเป็นเรื่องเดียวกัน โดยการเรียบเรียงและรวบรวม เพื่อคงไว้ซึ่งแบบแผนหรือโครงสร้างใหม่ แบ่งออกเป็น 5.10 ผลผลิตจากการสื่อความหมายเฉพาะ เป็นการสื่อความคิด ความรู้สึก และ/หรือประสบการณ์ของผู้เขียนหรือผู้พูดไปยังผู้อื่น 5.20 ผลผลิตจากแผนงานหรือชุดเสนอเพื่อปฏิบัติการ เป็นการพัฒนาแผนการทำงานหรือโครงสร้างของแผนปฏิบัติการ โดยแผนควรสอดคล้องกับความต้องการที่ต้องการให้ผู้เรียนพัฒนาตนเอง 5.30 การได้มาซึ่งชุดของความสัมพันธ์เชิงนามธรรม เป็นการพัฒนาชุดของความสัมพันธ์เชิงนามธรรม เพื่อจัดจำแนกหรือเพื่ออธิบายข้อมูลหรือปรากฏการณ์บางอย่าง หรือการอนุมานข้อความและความสัมพันธ์จากชุดของข้อความเบื้องต้นหรือการนำเสนอด้วยสัญลักษณ์

6.00 การประเมิน (Evaluation) การตัดสินเกี่ยวกับคุณค่าของเนื้อหาและวิธีการ ตามจุดมุ่งหมายที่กำหนด การตัดสินเชิงปริมาณและคุณภาพภายใต้ขอบเขตของเนื้อหาและวิธีการให้สอดคล้องกับเกณฑ์ หรือมาตรฐาน แบ่งออกเป็น 6.10 การตัดสินตามเกณฑ์ภายใน เป็นการประเมินความถูกต้องของการสื่อความหมายจากหลักฐาน ตามตรรกวิทยา ความสอดคล้อง และเกณฑ์ภายใน 6.20 การตัดสินตามเกณฑ์ภายนอก เป็นการประเมินวัตถุโดยอาศัยเกณฑ์ที่เลือกหรือจํา

### **ทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory: CTT)**

ศิริชัย กาญจนวาสี (2544) กล่าวถึงจุดมุ่งหมายของทฤษฎีการทดสอบแบบดั้งเดิมไว้ว่า เพื่อตรวจสอบความสัมพันธ์ระหว่างคะแนนสังเกตได้กับคะแนนที่แท้จริง และวิเคราะห์คุณภาพโดยส่วนรวมของข้อสอบและแบบสอบ ที่ใช้กับแต่ละกลุ่มคนในสถานการณ์สอบเฉพาะ โดยมีข้อตกลงเบื้องต้นที่สำคัญว่า ความคลาดเคลื่อนจากการวัดมีแบบแผนคงที่เหมือนกันสำหรับทุกกลุ่มที่ตอบข้อสอบ จึงไม่สนใจศึกษาพฤติกรรมการตอบข้อสอบของแต่ละบุคคลเป็นรายข้อ แต่ต้องการสภาพรวมในการตอบข้อสอบของกลุ่มบุคคลนั้นโดยเฉพาะ

## ทฤษฎีความเที่ยง (Theory of Reliability)

ศิริชัย กาญจนวาสี (2544) กล่าวว่า ความเที่ยง (Reliability) เป็นอัตราส่วนระหว่างความแปรปรวนของคะแนนจริงกับความแปรปรวนของคะแนนที่สังเกตได้ หรือเป็นค่ากำลังสองของสหสัมพันธ์ระหว่างคะแนนจริงกับคะแนนที่สังเกตได้ ความเที่ยง จึงหมายถึง ความคงเส้นคงวาของคะแนนสอบ แบ่งเป็น 4 ประเภท คือ 1) ความเที่ยงแบบคงที่สามารถประมาณค่าได้โดยวิธีสอบซ้ำด้วยแบบสอบเดิม 2) ความเที่ยงแบบสมมูล สามารถประมาณค่าได้โดยวิธีใช้แบบสอบสมมูลสอบในช่วงเวลาเดียวกัน 3) ความเที่ยงแบบคงที่และสมมูล สามารถประมาณค่าได้โดยวิธีสอบซ้ำด้วยแบบสอบที่สมมูลกัน และ 4) ความเที่ยงแบบความสอดคล้องภายใน ประมาณค่าได้โดยวิธีแบ่งครึ่งข้อสอบ วิธีของคูเดอร์-ริชาร์ดสัน วิธีสัมประสิทธิ์แอลฟาของครอนบาค และวิธีวิเคราะห์ความแปรปรวนของฮอยท์ ซึ่งวิธีการประมาณค่าที่ต่างกันให้ผลการประมาณค่าต่างกัน

ปัจจัยสำคัญที่มีผลต่อค่าสัมประสิทธิ์ความเที่ยงของแบบสอบ ได้แก่ 1) ความเป็นเอกพันธ์ของกลุ่มผู้สอบ (Group Homogeneity) ถ้าควบคุมปัจจัยอื่นให้คงที่ กลุ่มผู้สอบที่มีความสามารถจริงยิ่งใกล้เคียงกันมากเท่าใด ค่าสัมประสิทธิ์ความเที่ยงของแบบสอบยิ่งต่ำลงเท่านั้น 2) ความยาวของแบบสอบ เมื่อความยาวของแบบสอบเพิ่มขึ้น ความเที่ยงของแบบสอบจะเพิ่มขึ้นด้วย 3) ความสัมพันธ์ระหว่างข้อสอบ (Interitem Correlation) ข้อสอบที่มีความสัมพันธ์กันสูง ค่าสัมประสิทธิ์ความเที่ยงของแบบสอบจะสูง เนื่องจากเนื้อหาเป็นเอกพันธ์ (Guilford, 1954 อ้างใน ศิริชัย กาญจนวาสี, 2544: 64) 4) เวลาที่ใช้ในการทำแบบสอบ (Time Limit) การกำหนดเวลาสอบมากหรือน้อยเกินไป มีผลให้ความเที่ยงของแบบสอบลดลง 5) วิธีที่ใช้ในการประมาณค่าความเที่ยง

ความสัมพันธ์ระหว่างความยากง่ายและอำนาจจำแนกของข้อสอบ พบว่า ข้อสอบที่มีค่าความยากเกินไปหรือง่ายเกินไปจะไม่สามารถจำแนกผู้เรียนเก่งกับอ่อนออกจากกันได้ หรือจำแนกผู้เรียนที่มีความสามารถพิเศษกับที่ไม่มีความสามารถพิเศษออกจากกันได้ และสามารถสรุปความสัมพันธ์ได้อีกดังนี้ (ศิริชัย กาญจนวาสี, 2544: 188) 1) ช่วงของค่าอำนาจจำแนกของข้อสอบขึ้นอยู่กับค่าความยากง่ายของข้อสอบ 2) ข้อสอบที่ง่ายมากหรือยากมาก จะมีค่าอำนาจจำแนกอยู่ในช่วงจำกัดและต่ำ 3) ข้อสอบที่มีความยากง่ายปานกลาง อาจมีค่าอำนาจจำแนกเป็นบวก ลบ หรือ ศูนย์ก็ได้ ขึ้นกับสัดส่วนของผู้ตอบข้อสอบข้อนั้นถูกในกลุ่มสูงและกลุ่มต่ำ 4) ค่าอำนาจจำแนกของข้อสอบสูงสุดเท่ากับ 1.00 หรือต่ำสุดเท่ากับ -1.00 เป็นไปได้สำหรับข้อสอบที่มีค่าความยากง่ายเท่ากับ 0.50 เท่านั้น

Karelia et al. (2013) ศึกษาระดับความยากง่ายและอำนาจจำแนก และความสัมพันธ์ระหว่างค่าความยากง่ายและอำนาจจำแนกของข้อสอบแบบเลือกตอบ 4 ตัวเลือก ของแบบสอบวิชาเภสัชวิทยา ผลการศึกษาพบว่า ร้อยละ 29 ของข้อสอบทั้งหมด เป็นข้อสอบง่าย ร้อยละ 78



ของข้อสอบทั้งหมด มีค่าอำนาจจำแนก ใช้ได้ ร้อยละ 46 ของข้อสอบทั้งหมด มีอำนาจจำแนกดีเยี่ยม  
อำนาจจำแนกสัมพันธ์กับค่าความยากง่ายทางบวกในระดับต่ำ อย่างไม่มีนัยสำคัญทางสถิติที่ระดับ  
0.05

Pande et al. (2013) ศึกษาความสัมพันธ์ระหว่างดัชนีความยากง่ายและอำนาจจำแนก  
ของข้อสอบแบบเลือกตอบในการสอบเพื่อติดตามความก้าวหน้าวิชาสรีรวิทยา ผลการศึกษาพบว่า  
ข้อสอบ ส่วนใหญ่ ร้อยละ 75 มีดัชนีความยากง่ายและการจำแนกใช้ได้ ข้อสอบที่มีระดับความ  
ยากง่ายปานกลางจะมีความสามารถในการจำแนกสูงที่สุด ข้อสอบที่ง่ายเกินไปหรือยากเกินไป จะมี  
ความสามารถในการจำแนกต่ำ ร้อยละ 4 ของข้อสอบทั้งหมด มีค่าอำนาจจำแนกเป็นลบ แสดงว่า  
ข้อสอบผิดหรือเฉลยไม่ถูกต้อง

Hotiu (2006) ศึกษาความสัมพันธ์ระหว่างค่าความยากง่ายและอำนาจจำแนกของข้อสอบ  
แบบเลือกตอบในหลักสูตรวิทยาศาสตร์กายภาพ ผลการศึกษาพบว่า เมื่อความยากง่ายเพิ่มขึ้น  
ข้อสอบสามารถจำแนกผู้เรียนที่มีความสามารถต่างกันได้เพิ่มขึ้น เมื่อข้อสอบยากเกินไปหรือง่ายเกินไป  
ความสามารถของข้อสอบในการจำแนกผู้เรียนจะลดลง ข้อเสนอแนะควรพัฒนาข้อสอบที่มีอำนาจ  
จำแนกเหมาะสม

Sim & Rasiah (2006) ศึกษาความสัมพันธ์ระหว่างดัชนีความยากและอำนาจจำแนก  
ในข้อสอบหลายตัวเลือกชนิดถูก-ผิดของข้อสอบของสหสาขาทางคลินิก ผลการศึกษาพบว่า ข้อสอบ  
แบบเลือกตอบที่มีอำนาจจำแนกดีมีแนวโน้มว่าข้อสอบจะมีความยากง่ายในระดับปานกลาง และ  
ข้อสอบที่มีความยากง่ายปานกลางถึงยากเกินไปมีแนวโน้มจะให้อำนาจจำแนกทางลบ

ในปีการศึกษา 2556 นักศึกษาพยาบาลตำรวจทั้ง 4 ชั้นปี ได้ใช้หลักสูตรทั้ง 2 หลักสูตร คือ  
ชั้นปีที่ 1 และชั้นปีที่ 2 ใช้หลักสูตรพยาบาลศาสตรบัณฑิต หลักสูตรปรับปรุง พ.ศ. 2554 ส่วนชั้น  
ปีที่ 3 และชั้นปีที่ 4 ใช้หลักสูตรพยาบาลศาสตรบัณฑิต ฉบับปรับปรุง พ.ศ. 2549 รวม 19 รายวิชา

### สมมติฐานการวิจัย

1. ความสัมพันธ์ระหว่างดัชนีความยากง่ายและอำนาจจำแนกของข้อสอบ ไม่เป็นเชิงเส้นตรง
2. ขนาดความสัมพันธ์เชิงเส้นตรงระหว่างดัชนีความยากง่ายและอำนาจจำแนกของข้อสอบ  
อยู่ในระดับต่ำ
3. ทิศทางความสัมพันธ์ระหว่างดัชนีความยากง่ายและอำนาจจำแนกของข้อสอบเป็นบวก



## วิธีดำเนินการวิจัย

ข้อมูล ได้แก่ ข้อสอบ และผลการวิเคราะห์ข้อสอบ ได้แก่ ความยากง่าย อำนาจจำแนก และประสิทธิภาพตัวลวงของข้อสอบปลายภาคแบบเลือกตอบ กลุ่มวิชาชีพ ภาคทฤษฎี หลักสูตรพยาบาลศาสตรบัณฑิต วิทยาลัยพยาบาลตำรวจ ปีการศึกษา 2556 จำนวน 19 ฉบับหรือ 19 รายวิชา

เครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูล ได้แก่ แบบบันทึกข้อมูลเบื้องต้นของข้อสอบ ประกอบด้วย รายวิชาที่ จำนวนผู้สอบ คะแนนต่ำสุด-สูงสุด คะแนนเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน และ KR-20 และแบบบันทึกผลการวิเคราะห์ข้อสอบ ประกอบด้วยจุดมุ่งหมายทางการศึกษา ค่าความยากง่าย ค่าอำนาจจำแนกของตัวเลือกที่ถูกต้อง และตัวลวง

เกณฑ์การบันทึกจุดมุ่งหมายการวัดของข้อสอบ ตามสารบบจำแนกของบลูม (Bloom, 1956 อ้างใน Reilly & Oermann, 1999) แบ่งออกเป็น 6 ชั้น ได้แก่ ความรู้ความจำ ความเข้าใจ การประยุกต์/ การนำไปใช้ การวิเคราะห์ การสังเคราะห์ และการประเมินค่า

เกณฑ์ที่ใช้ในการตัดสินความยากง่ายของข้อสอบ แบ่งเป็น 5 ระดับ คือ (ศิริชัย กาญจนวาสี, 2544) ง่ายมาก หมายถึง ค่า  $P = 0.80-1.00$  ค่อนข้างง่าย หมายถึง ค่า  $P = 0.60-0.79$  ปานกลาง หมายถึง ค่า  $P = 0.40-0.59$  ค่อนข้างยาก หมายถึง ค่า  $P = 0.20-0.39$  และยากมาก หมายถึง ค่า  $P = 0.00-0.19$

เกณฑ์ที่ใช้ในการตัดสินอำนาจจำแนกของข้อสอบ แบ่งเป็น 5 ระดับ คือ (ศิริชัย กาญจนวาสี, 2544) ดีมาก หมายถึง ค่า  $r = 0.60-1.00$  ดี หมายถึง ค่า  $r = 0.40-0.59$  พอใช้ได้ หมายถึง ค่า  $r = 0.20-0.39$  ค่อนข้างต่ำ ควรปรับปรุง หมายถึง ค่า  $r = 0.10-0.19$  ต่ำมาก ควรปรับปรุง หมายถึง ค่า  $r = 0.00-0.09$

เกณฑ์ที่ใช้ในการตัดสิน ความยากง่ายและอำนาจจำแนกของตัวลวง มีค่า  $P$  และ  $r$  ของตัวลวง ตั้งแต่ 0.05 ขึ้นไป

เกณฑ์การแปลค่าความสัมพันธ์ (Burns & Bush, 2003: 534)  $\pm 0.81$  ถึง  $\pm 1.00$  หมายถึง สูง  $\pm 0.61$  ถึง  $\pm 0.80$  หมายถึง ปานกลาง  $\pm 0.41$  ถึง  $\pm 0.60$  หมายถึง ต่ำ  $\pm 0.21$  ถึง  $\pm 0.40$  หมายถึง ต่ำมาก  $\pm 0.00$  ถึง  $\pm 0.20$  ไม่สัมพันธ์กัน

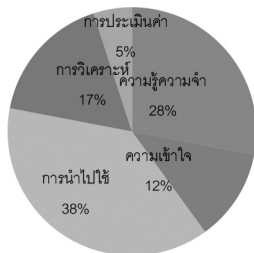
การเก็บรวบรวมข้อมูล ได้แก่ 1) ขออนุญาตผู้บังคับการวิทยาลัยพยาบาลตำรวจ 2) ขออนุญาตคณะกรรมการจริยธรรมและวิจัย ของโรงพยาบาลตำรวจ 3) ขอความร่วมมือจากภาคีวิชาการในการส่งข้อสอบ และผลการวิเคราะห์ข้อสอบ 4) บันทึกข้อมูลในแบบบันทึก 5) เอกสารที่เกี่ยวข้องทุกอย่าง

จะถูกเก็บเป็นความลับ การนำเสนอผลการวิเคราะห์ข้อมูลจะนำเสนอเป็นรายวิชา และในรายงานวิจัยฉบับสมบูรณ์จะไม่นำเสนอชื่อรายวิชา และ 6) หลังดำเนินการเสร็จเรียบร้อยแล้ว คณะผู้วิจัยจะส่งเอกสารทั้งหมดคืนไปยังกลุ่มงานอาจารย์

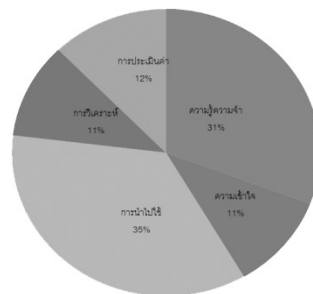
สถิติบรรยายใช้วิเคราะห์ข้อมูลเบื้องต้นของข้อสอบ และวิเคราะห์ความสัมพันธ์ระหว่างค่าความยากง่าย และค่าอำนาจจำแนกของข้อสอบ ด้วยการวิเคราะห์ค่าสหสัมพันธ์แบบเพียร์สัน โดยใช้โปรแกรมสำเร็จรูป ค่า  $p\text{-value} < 0.05$  พิจารณาว่ามีความสำคัญทางสถิติ และนำเสนอข้อมูลผลการศึกษาในรูปแบบตาราง กราฟ และแผนภาพกระจาย (scatter plot)

## ผลการวิจัย

1. ระดับการวัดตามจุดมุ่งหมายทางการศึกษาด้านพุทธิปริเขต มีเพียง 5 ระดับ ได้แก่ ความรู้ความจำ ความเข้าใจ การนำไปใช้ การวิเคราะห์ และการประเมินค่า จำนวน 1 รายวิชา วัดระดับความรู้ความจำเพียงระดับเดียว จำนวน 1 รายวิชา วัด 2 ระดับ ส่วนใหญ่วัดระดับความรู้ความจำ และอีกส่วนหนึ่งวัดระดับความเข้าใจ จำนวน 3 รายวิชาที่วัด 3 ระดับ ส่วนใหญ่วัดความรู้ความจำ อีก 2 ระดับ ได้แก่ ความเข้าใจและการนำไปใช้ จำนวน 6 รายวิชา วัด 4 ระดับ ได้แก่ ความรู้ความจำ ความเข้าใจ การนำไปใช้ และการวิเคราะห์ และมีจำนวน 2 รายวิชาที่วัด 5 ระดับ ได้แก่ รายวิชาที่ 9 และ 13 ซึ่งพบว่า 1 ใน 3 เน้นการวัดความคิดระดับสูง หรือวัดระดับการนำไปใช้ขึ้นไป ส่วนรายวิชาที่ 13 พบว่า ร้อยละ 58 เน้นการวัดความคิดระดับสูง คือวัดระดับการนำไปใช้ขึ้นไป ดังแสดงในภาพ 1 และ 2



ภาพ 1 ผลการวิเคราะห์จุดมุ่งหมายทางการศึกษาด้านพุทธิปริเขตของข้อสอบรายวิชาที่ 9

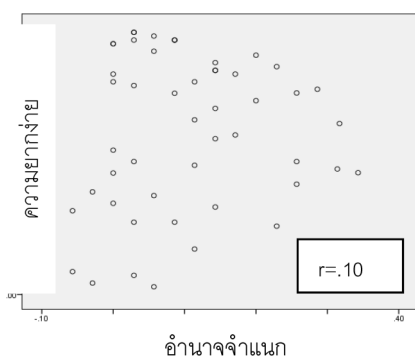


ภาพ 2 ผลการวิเคราะห์จุดมุ่งหมายทางการศึกษาด้านพุทธิปริเขตของข้อสอบรายวิชาที่ 13

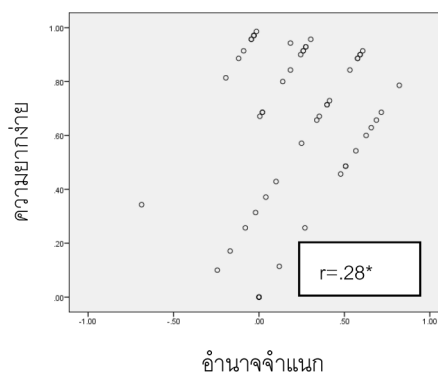
2. ค่าเฉลี่ยของดัชนีความยากง่ายของข้อสอบทั้ง 19 ฉบับ อยู่ในเกณฑ์ปานกลาง (.45) ถึงค่อนข้างง่าย (.77) ส่วนค่าเฉลี่ยของอำนาจจำแนกของข้อสอบทั้ง 19 ฉบับ อยู่ในเกณฑ์ต่ำมาก ควรปรับปรุง (.08) ถึง พอใช้ได้ (.31) รายวิชาที่แบบสอบมีข้อสอบที่ดี คือ ดัชนีความยากง่ายและดัชนีอำนาจจำแนกอยู่ในเกณฑ์ที่ยอมรับได้พร้อมกันมากที่สุด ได้แก่ รายวิชาที่ 9 คิดเป็นร้อยละ 40.00

3. ตัวลวงที่ดีทั้ง 3 ตัวเลือก พบเพียง 1-3 ข้อ คิดเป็นร้อยละ 1.67-10.00 ใน 10 รายวิชา

4. ไม่พบความสัมพันธ์เชิงเส้นตรงระหว่างดัชนีความยากง่ายและอำนาจจำแนกของข้อสอบจำนวน 16 จาก 19 ฉบับ และทิศทางความสัมพันธ์ระหว่างดัชนีความยากง่ายและอำนาจจำแนกของข้อสอบเป็นบวก ใน 15 จาก 19 ฉบับ ตัวอย่างไม่พบความสัมพันธ์เชิงเส้นตรงแสดงในภาพ 3 และพบความสัมพันธ์เชิงเส้นตรงแสดงในภาพ 4



**ภาพ 3** ความสัมพันธ์ระหว่างดัชนีความยากง่ายและอำนาจจำแนกของข้อสอบในรายวิชาที่ 1



**ภาพ 4** ความสัมพันธ์ระหว่างดัชนีความยากง่ายและอำนาจจำแนกของข้อสอบในรายวิชาที่ 19

## อภิปรายผลการวิจัย

1. จุดมุ่งหมายทางการศึกษาส่วนใหญ่ของข้อสอบแบบเลือกตอบปลายภาค 13 ฉบับใน 13 รายวิชา กลุ่มวิชาชีพ หลักสูตรพยาบาลศาสตรบัณฑิต วิทยาลัยพยาบาลตำรวจ ปีการศึกษา 2556 ส่วนใหญ่อยู่ในระดับการวัดความรู้ความจำมากที่สุด รองลงมาคือ การนำไปใช้ ความเข้าใจ การวิเคราะห์ และการประเมินค่า ตามลำดับ เนื่องจากข้อจำกัดของการเขียนข้อสอบแบบเลือกตอบก็คือ สร้างยาก มีกระบวนการสร้างที่ซับซ้อนและใช้เวลาในการสร้างมาก (Karelia et al., 2013: 41) ทำให้ผู้ออกข้อสอบมักจะออกข้อสอบโดยการคัดลอกเนื้อหาจากตำรา เพราะเขียนได้ง่าย ซึ่งทำให้วัดในระดับความรู้ ความจำ เป็นส่วนใหญ่และละเลยการเขียนข้อสอบที่วัดความคิดระดับสูง

2. ดัชนีความยากง่ายของข้อสอบทั้ง 19 รายวิชา อยู่ในเกณฑ์ปานกลางถึงค่อนข้างง่าย สอดคล้องกับผลการวิเคราะห์จุดมุ่งหมายทางการศึกษาที่พบว่า ข้อสอบส่วนใหญ่วัดในระดับความรู้ ความจำ สำหรับอำนาจจำแนกของข้อสอบทั้ง 19 รายวิชา อยู่ในเกณฑ์ต่ำมาก ควรปรับปรุง ถึงพอใช้ได้ และเมื่อพิจารณาทั้งดัชนีความยากง่ายและอำนาจจำแนกของข้อสอบอยู่ในเกณฑ์ที่ยอมรับได้พร้อมกัน พบว่าข้อสอบแต่ละรายวิชามีข้อสอบที่ดีต่ำกว่าร้อยละ 50.00 ส่วนใหญ่มีดัชนีอำนาจจำแนกติดลบ ซึ่งไม่สามารถจำแนกผู้สอบที่เก่งออกจากผู้สอบที่อ่อนได้ โดยพบในทุกรายวิชา Pande et al. (2013) กล่าวว่า ข้อสอบที่มีอำนาจจำแนกเป็นลบ แสดงว่า ข้อสอบผิดหรือเฉลย ไม่ถูกต้อง หรือผู้สอบเลือกตัวลวงเป็นคำตอบมากกว่าเลือกตัวเลือกที่ถูกต้อง ซึ่งสอดคล้องกับ ศิริชัย กาญจนาวาสี (2544) กล่าวว่า ช่วงของอำนาจจำแนกของข้อสอบขึ้นอยู่กับค่าความยากง่ายของ ข้อสอบ เมื่อข้อสอบส่วนใหญ่ในการศึกษาครั้งนี้ มีความยากง่ายอยู่ในระดับปานกลางถึงค่อนข้างง่าย จึงทำให้ค่าอำนาจจำแนกของข้อสอบต่ำ และไม่พบว่าข้อสอบในรายวิชาใดมีสัดส่วนของข้อสอบที่มีความยากง่ายเป็นไปตามที่ EXHCOBA จึงแนะนำให้แบ่งข้อสอบเป็น ง่ายมาก ร้อยละ 5 ค่อนข้าง ง่าย ร้อยละ 20 ปานกลาง ร้อยละ 50 ค่อนข้างยาก ร้อยละ 20 และยากมากร้อยละ 5

3. ผลการศึกษานี้ เป็นไปตามสมมติฐานการวิจัยที่ว่า ความสัมพันธ์ระหว่างดัชนีความยากง่าย และอำนาจจำแนกของข้อสอบ ไม่เป็นลักษณะเชิงเส้นตรง คือ ไม่มีนัยสำคัญทางสถิติ จำนวน 15 จาก 19 ฉบับ ยกเว้น รายวิชาที่ 9, 11, 12 และ 19 เมื่อพิจารณาข้อสอบรายข้อในรายวิชาที่ 9, 11, 12 และ 19พบว่า เมื่อข้อสอบค่อนข้างง่ายถึงง่ายมาก ความสามารถในการจำแนกผู้สอบที่เก่ง ออกจากผู้สอบที่อ่อน จะอยู่ในระดับค่อนข้างต่ำถึงพอใช้ได้ และข้อสอบที่ค่อนข้างยากถึงยากมาก จะมีค่าอำนาจจำแนกต่ำมาก หรือค่าอำนาจจำแนกติดลบ จึงทำให้พบว่าดัชนีความยากง่ายมีความสัมพันธ์ทางบวกกับดัชนีอำนาจจำแนกอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 โดยพบว่ามีความสัมพันธ์ อยู่ในระดับต่ำ สอดคล้องกับการศึกษาของ Karelia et al. (2013) และ Pande et al. (2013) และ Sim & Rasiah (2006) ที่พบอำนาจจำแนกสัมพันธ์กับค่าความยากง่ายทางบวกในระดับต่ำ อย่างไม่มีนัยสำคัญทางสถิติ และความสัมพันธ์ไม่เป็นลักษณะเชิงเส้นตรง

4. ผลการศึกษานี้ สอดคล้องกับสมมติฐานที่ว่าทิศทางความสัมพันธ์ระหว่างดัชนีความยากง่าย และอำนาจจำแนกของข้อสอบเป็นบวกใน 15 ฉบับ ยกเว้นในรายวิชาที่ 3, 4, 6 และ 15 ซึ่งพบ ความสัมพันธ์ทางลบ แต่ไม่มีนัยสำคัญทางสถิติ เมื่อพิจารณาดัชนีความยากง่ายและอำนาจจำแนก ของข้อสอบในรายวิชา 3, 4, 6 และ 15 พบว่า รายวิชาที่ 3 ร้อยละ 54.00 อยู่ในเกณฑ์ง่ายมาก แต่อำนาจจำแนกต่ำมาก ควรปรับปรุง ร้อยละ 46 รายวิชาที่ 4 อยู่ในเกณฑ์ยากมากและง่ายมาก ร้อยละ 30 เท่ากัน แต่อำนาจจำแนกต่ำมาก ควรปรับปรุง ร้อยละ 40 รายวิชาที่ 6 อยู่ในเกณฑ์

ง่ายมาก ร้อยละ 31 แต่อำนาจจำแนกต่ำมาก ควรปรับปรุง ร้อยละ 43 และรายวิชาที่ 15 ร้อยละ 67.00 อยู่ในเกณฑ์ค่อนข้างง่าย แต่อำนาจจำแนกต่ำมาก ควรปรับปรุง ร้อยละ 43 ดังนั้น จึงพบความสัมพันธ์ทางลบระหว่างดัชนีความง่ายและดัชนีอำนาจจำแนกของข้อสอบ เพราะเมื่อข้อสอบยากมากหรือง่ายมาก อำนาจในการจำแนกผู้สอบที่มีความสามารถสูงออกจากผู้สอบที่มีความสามารถต่ำจะลดลง (ศิริชัย กาญจนวาสี, 2544)

### ข้อเสนอแนะ

ข้อเสนอแนะในการนำผลการวิจัยไปใช้ ได้แก่ 1) ผู้รับผิดชอบรายวิชาดำเนินการประชุมชี้แจงผลที่สะท้อนผลการจัดการเรียนการสอน และการปรับปรุงคุณภาพข้อสอบให้มีมาตรฐาน 2) ผู้บริหารวางแผนการเพิ่มคุณภาพของข้อสอบ และสื่อสารไปยังผู้เกี่ยวข้อง เพื่อให้การวัดและประเมินผลการศึกษาที่มีคุณภาพสูงขึ้น และ 3) ผู้สอนนำจุดอ่อนหรือข้อที่ควรปรับปรุงไปวางแผนพัฒนาเครื่องมือในการประเมินผลการเรียนรู้ของผู้เรียนให้มีคุณภาพ

ข้อเสนอแนะในการวิจัยต่อไป 1) พัฒนาแบบสอบมาตรฐานในรายวิชาทางการพยาบาล 2) พัฒนาดังข้อสอบในรายวิชาทางการพยาบาล และ 3) พัฒนาโปรแกรมการเขียนข้อสอบวัดทักษะการคิดระดับสูงสำหรับผู้สอน

### เอกสารอ้างอิง

- วิทยาลัยพยาบาลตำรวจ โรงพยาบาลตำรวจ สำนักงานตำรวจแห่งชาติ. (2549). *หลักสูตรพยาบาล-ศาสตร์บัณฑิต ฉบับปรับปรุง พ.ศ. 2549*. เอกสารอัดสำเนา.
- วิทยาลัยพยาบาลตำรวจ โรงพยาบาลตำรวจ สำนักงานตำรวจแห่งชาติ. (2554). *หลักสูตรพยาบาล-ศาสตร์บัณฑิต หลักสูตรปรับปรุง พ.ศ. 2554*. เอกสารอัดสำเนา.
- ศิริชัย กาญจนวาสี. (2544). *ทฤษฎีการทดสอบแบบดั้งเดิม*. (พิมพ์ครั้งที่ 4). กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- อุทุมพร ทองอุไทย (ผู้แปล). (2523). *สารบบจำแนกจุดมุ่งหมายทางการศึกษา: การจัดทำพวกวัตถุประสงค์ทางการศึกษา คู่มือเล่ม 1: พุทธิปริเขต*. โดย เบนจามิน เอส บลูม บรรณาธิการ แมกซ์ ดี เองเกิลฮาร์ท เอ็ดเวิร์ด เจ. เฟิสส์ท วอลท์เกอร์ เอช. ฮิลล์ และเดวิด อาร์. แครธไวล. กรุงเทพฯ: โรงพิมพ์จุฬาลงกรณ์มหาวิทยาลัย.
- Burns, A. C., & Bush, R. F. (2003). *Marketing research: Online research applications* (4<sup>th</sup> ed). Pearson Prentice Hall.

- Escudero, E. B., Reyna, N. L., & Morales, M. R. (2000). The level of difficulty and discrimination power of the basic knowledge and skills examination (EXHCOBA). *Revista Electronica de Investigacion Educativa*, 2(1), 1-16.
- Hotiu, A. (2006). *The relationship between item difficulty and discrimination indices in Multiple-Choice Tests in a physical science course*. A thesis is submitted to the faculty of Charles Schmidt College of Science in partial fulfillment of the requirements for the degree of master in science, Florida Atlantic University.
- Karelia, B. N., Pillai, A., & Vegada, V. N. (2013). The levels of difficulty and discrimination indices and relationship between them in four-response type multiple choice questions of pharmacology summative tests of Year II M.B.B.S students. *International e-Journal of Science, Medicine & Education*, 7(2), 41-46.
- Mitra, N. K., Nagaraja, H. S., Ponnudurai, G., & Judson, J. P. (2009). The levels of difficulty and discrimination indices in type A multiple choice questions of pre-clinical semester 1 multidisciplinary summative tests. *International e-Journal of Science, Medicine & Education*, 3(1), 2-7.
- Pande, S. S., Pande, S. R., Parate, V. R., Nikam, A. P., & Agrekar, S. H. (2013). Correlation between difficulty & discrimination indices of MCQs in formative exam in physiology. *South-East Asian Journal of Medical Education*, 7(1),45-50.
- Reilly, D. E., & Oerman, M. H. (1999). *Behavioral objectives: Evaluation in nursing*. (3<sup>rd</sup> ed). New York: National League for Nursing.
- Sabri, S. (2013). Item analysis of student comprehensive test for research in teaching beginner string ensemble using model based teaching among music students in public universities. *International Journal of Education and Research*, 1(12), 1-14.
- Sim, S-M., & Rasiah, R. I. (2006). Relationship between item difficulty and discrimination indices in True/False type multiple choice questions of a Para-clinical multidisciplinary paper. 2006. *Annals Academy of Medicine*, 35(2), 67-71.